

## Rochester Institute of Technology RIT Scholar Works

---

Theses

Thesis/Dissertation Collections

---

8-1-2008

# Face recognition in low resolution video sequences using super resolution

Somi Ruwan Arachchige

Follow this and additional works at: <http://scholarworks.rit.edu/theses>

---

### Recommended Citation

Arachchige, Somi Ruwan, "Face recognition in low resolution video sequences using super resolution" (2008). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by the Thesis/Dissertation Collections at RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact [ritscholarworks@rit.edu](mailto:ritscholarworks@rit.edu).

# Face Recognition in Low Resolution Video Sequences using Super Resolution

by

Somi Ruwan Budhagoda Arachchige

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of  
Master of Science in Computer Engineering

Supervised by

Professor Dr. Andreas Savakis  
Department of Computer Engineering  
Kate Gleason College of Engineering  
Rochester Institute of Technology  
Rochester, NY  
August 2008

**Approved By:**

---

**Dr. Andreas Savakis** - Professor and Department Head  
*Primary Advisor – R.I.T. Dept. of Computer Engineering*

---

**Dr. Shanchieh Jay Yang** - Assistant Professor  
*Secondary Advisor – R.I.T. Dept. of Computer Engineering*

---

**Dr. Marcin Lukowiak** - Assistant Professor  
*Secondary Advisor – R.I.T. Dept. of Computer Engineering*

# **Thesis Release Permission Form**

**Rochester Institute of Technology  
Kate Gleason College of Engineering**

**Title: Face Recognition in Low Resolution Video Sequences using Super Resolution**

I, Somi Ruwan Budhagoda Arachchige, hereby grant permission to the Wallace Memorial Library to reproduce my thesis in whole or part.

---

Somi Ruwan Budhagoda Arachchige

---

Date

# **Dedication**

I would like to dedicate this work to my Mom, for all her encouragement, support, and wisdom. Also, to the memories of my belloved late father Piyasiri Budhagoda Arachchige.

# Acknowledgments

Fist and foremost, I would like to thank my advisor Dr. Andreas Savakis serving as a patient mentor throughout this thesis and for his faith in my abilities. His constant interest helped at every stage to enable this thesis to be completed.

Also, I would like to thank my committee, Dr. Shanchieh Yang and Dr. Marcin Lukowiak and Computer Engineering faculty for their support.

I would specially like to thank my family, uncle Wicky and Champika Gunasekera for encouraging and supporting me throughout the thesis

## **Abstract**

Human activity is a major concern in a wide variety of applications, such as video surveillance, human computer interface and face image database management. Detecting and recognizing faces is a crucial step in these applications. Furthermore, major advancements and initiatives in security applications in the past years have propelled face recognition technology into the spotlight.

The performance of existing face recognition systems declines significantly if the resolution of the face image falls below a certain level. This is especially critical in surveillance imagery where often, due to many reasons, only low-resolution video of faces is available. If these low-resolution images are passed to a face recognition system, the performance is usually unacceptable. Hence, resolution plays a key role in face recognition systems. In this thesis, we address this issue by using super-resolution techniques as a middle step, where multiple low resolution face image frames are used to obtain a high-resolution face image for improved recognition rates. Two different techniques based on frequency and spatial domains were utilized in super resolution image enhancement. In this thesis, we apply super resolution to both images and video utilizing these techniques and we employ principal component analysis for face matching, which is both computationally efficient and accurate. The result is a system that can accurately recognize faces using multiple low resolution images/frames.

# Table of Contents

	Dedication	.....	iii
	Acknowledgements	.....	iv
	Abstract	.....	v
	Table of Contents	.....	vi
	List of Figures	.....	vii
	List of Tables	.....	x
	Glossary	.....	xi
<b>1.0</b>	<b>Introduction</b>	.....	1
1.1	<i>Motivation</i>	.....	2
1.2	<i>Problem Statement</i>	.....	2
1.3	<i>Outline</i>	.....	5
<b>2.0</b>	<b>Background</b>	.....	6
2.1	<i>Face Detection &amp; Recognition</i>	.....	7
2.2	<i>Super Resolution Imaging</i>	.....	9
2.3	<i>Application Domains</i>	.....	12
<b>3.0</b>	<b>Theoretical Development</b>	.....	15
3.1	<i>Face Detection using Haar Like Features</i>	.....	15
3.2	<i>Super Resolution</i>	.....	19
3.2.1	Frequency Domain Approach	.....	19
3.2.1.1	<i>Rotation Estimation</i>	.....	21
3.2.1.2	<i>Shift Estimation</i>	.....	22
3.2.1.3	<i>Reconstruction</i>	.....	22
3.2.2	Spatial Domain Approach	.....	23
3.3	<i>Eigenface-based Facial Recognition</i>	.....	25
3.4	<i>Image and Video Databases</i>	.....	29
3.4.1	Data Preparation	.....	30
<b>4.0</b>	<b>System Development</b>	.....	33
4.1	<i>Face Detection</i>	.....	33
4.2	<i>Super Resolution</i>	.....	36
4.3	<i>Low Resolution Image Generator</i>	.....	39
4.4	<i>Face Recognition using Eigenfaces</i>	.....	40
4.4.1	Utilizing Eigenfaces	.....	42
4.5	<i>Limitations &amp; Difficulties</i>	.....	45
4.6	<i>Super Resolution GUI</i>	.....	47

<b>5.0</b>	<b>Evaluation</b>	.....	51
5.1	<i>Video Database Testing</i>	.....	51
5.2	<i>Image Database Testing</i>	.....	56
5.3	<i>Web Camera Captured Video Testing</i>	.....	62
5.4	<i>Timing Performance</i>	.....	66
5.5	<i>System Limitations</i>	.....	67
<b>6.0</b>	<b>Conclusion</b>	.....	68
6.1	<i>Future Work</i>	.....	69
	<b>Bibliography</b>	.....	70



## List of Figures

<b>Fig 1: (a)</b> Illustration of super resolution reconstruction; .....	4
<b>Fig.2:</b> General configuration of a face recognition system .....	8
<b>Fig 3:</b> Three basic steps of super resolution image reconstruction .....	10
<b>Fig 4:</b> Ideal super-resolution setup. ....	12
<b>Fig 5:</b> Some Haar-like features from [13] .....	16
<b>Fig 6:</b> The integral image trick. ....	17
<b>Fig 7:</b> The classifier cascade is a chain of single-feature filters.....	18
<b>Fig 8:</b> Graphical representation of super resolution problem.....	19
<b>Fig 9:</b> The amplitude of the Fourier transform of an image.....	21
<b>Fig 10:</b> 3 Eigenfaces .....	27
<b>Fig 11:</b> High-level functions of the Eigenface-based facial recognition algorithm .....	28
<b>Fig 12:</b> Original CVL database images and the resulting cropped images .....	30
<b>Fig 13:</b> Frame captures from four consecutive frames. ....	31
<b>Fig 14:</b> Steps of the Harr training process are shown above .....	34
<b>Fig 15: (a)</b> Camera captured high resolution frame .....	36
<b>Fig 16:</b> Generated images .....	37
<b>Fig 17:</b> (a) HR 90x120 image (b) four 45x60 generated low resolution images.....	40
<b>Fig 18:</b> First five Eigenfaces .....	42
<b>Fig 19:</b> Eigenfaces 60 & 80, 100 & 101, and 196.....	43
<b>Fig 20:</b> Relative magnitude of Eigenvalues .....	43

<b>Fig 21:</b> Number of Eigenfaces vs. face matching performance .....	44
<b>Fig 22:</b> Super resolution GUI .....	48
<b>Fig 23:</b> Input section of the super resolution GUI.....	49
<b>Fig 24:</b> Actions section of the super resolution GUI.....	49
<b>Fig 25:</b> Results GUI.....	50
<b>Fig 26:</b> Face matching results of video database with 15 subjects.....	54
<b>Fig 27:</b> Face matching results of VPA video database with 32 subjects.....	55
<b>Fig 28:</b> One subject from the Indian face database with two different poses.....	56
<b>Fig 29:</b> Face matching results of Indian face database with 55 subjects.....	57
<b>Fig 30:</b> Examples of unmatched images.....	58
<b>Fig 31:</b> Face matching results of CVL face database with 97 subjects. ....	59
<b>Fig 32:</b> SR results from two algorithms with areas selected .....	61
<b>Fig 33:</b> Captured & auto cropped frames from Logitech web camera .....	64
<b>Fig 34:</b> Face matching results of web camera captured video with 6 subjects. ....	65

## List of Tables

<b>Table 1a</b> - Plank video database first test results using LR 23x30 frames.....	52
<b>Table 1b</b> - Plank video database first test results using LR 45x60 frames.....	52
<b>Table 2a</b> - Plank video database results using LR 23x30 frames.....	53
<b>Table 2b</b> - Plank video database results using LR 45x60 frames.....	53
<b>Table 3a</b> - VPA video database results using LR 23x30 frames.....	54
<b>Table 3b</b> - VPA video database results using LR 45x60 frames.....	54
<b>Table 4a</b> - Indian face database using LR 23x30 images.....	57
<b>Table 4b</b> - Indian face database using LR 45x60 images.....	57
<b>Table 5a</b> - Results from the CVL image database using LR 23x30 images.....	59
<b>Table 5b</b> - Results from the CVL image database using LR 45x60 images.....	59
<b>Table 6</b> - Recognition performance (%) using cosine similarity.....	62
<b>Table 7a</b> - Face recognition results of web camera captured video for 23x30 frames....	64
<b>Table 7b</b> - Face recognition results of web camera captured video for 45x60 frames.....	64
<b>Table 8</b> - System timing performance.....	66

## Glossary

<b>Face Recognition</b>	Process of identifying the individual shown in a face image.
<b>Face Detection</b>	Determining the location and scale of a face in an image
<b>Face Matching</b>	Determining the degree of confidence to which a subject matches a facial image
<b>Eigenface</b>	An eigenvector from a set of face images
<b>GUI</b>	Graphical User Interface
<b>HR</b>	High Resolution
<b>LR</b>	Low Resolution
<b>SR</b>	Super Resolution
<b>PCA</b>	Principal Component Analysis
<b>SVM</b>	Support Vector Machine
<b>ROI</b>	Region of Interest

## Chapter 1 Introduction

Super resolution (SR) is a technique that aims to increase the resolution of an image or a sequence of images beyond the resolving power of the imaging system. Contrary to common belief, pixel count (i.e. the number of pixels in an image) is not an appropriate measure of image resolution. Instead, the resolution of an image should be defined as the amount of fine details that are visible. As a result, the industry practice of increasing the pixel count and reducing the pixel size does not always yield an improved resolution. This means that an image that is obtained by merely up sampling and interpolating a low resolution (LR) image does not have a higher resolution than its original. It has a larger number of pixels, but the resolving power remains the same; that is, the interpolated image does not contain more details. On the other hand, true resolution enhancement can be beneficial to a number of applications including face recognition.

Since the invention of computers, researchers have been trying to create systems with the capability to interpret data like humans. This includes tasks such as detection, localization, classification and discrimination of visual characteristics, as well as associating these features with contextual information. Even though many of the fundamental problems inherent in creating a system that can interpret data have been solved, there is still a great amount of work to be done. Face recognition has received significant attention during the last two decades and many researchers study various aspects of it. There are at least two reasons for this; the first one is a wide range of commercial and security applications and the second is the availability of feasible computer technology to develop and implement applications that demand strong computational power. Today, automatic recognition of human faces is a field

that gathers many researchers from different disciplines such as image processing, pattern recognition, computer vision, graphics, and psychology.

## **1.1 Motivation**

One of the key requirements for effective situation awareness is the acquisition of information at multiple scales [1]. While existing research has addressed several issues in the analysis of video, little work has been done in the area of better information acquisition of high resolution face images. Video sequences captured from modern digital cameras are normally digitized at lower resolutions than still pictures. A “freeze-frame” from such a video would thus be of lower than desirable quality. Therefore, resolution of images or video frames is a key factor in face recognition, since it is one of the limiting parameters in digital camera design. SR resolution is one of the key techniques exercised to improve low resolution images and could be used in low resolution surveillance video sequences to improve face recognition rates. Even though SR has been used extensively in many applications to improve the quality of still images, usage in video sequences has been minimal. Including SR in video opens possibilities for a lot of opportunities aside from standard video surveillance.

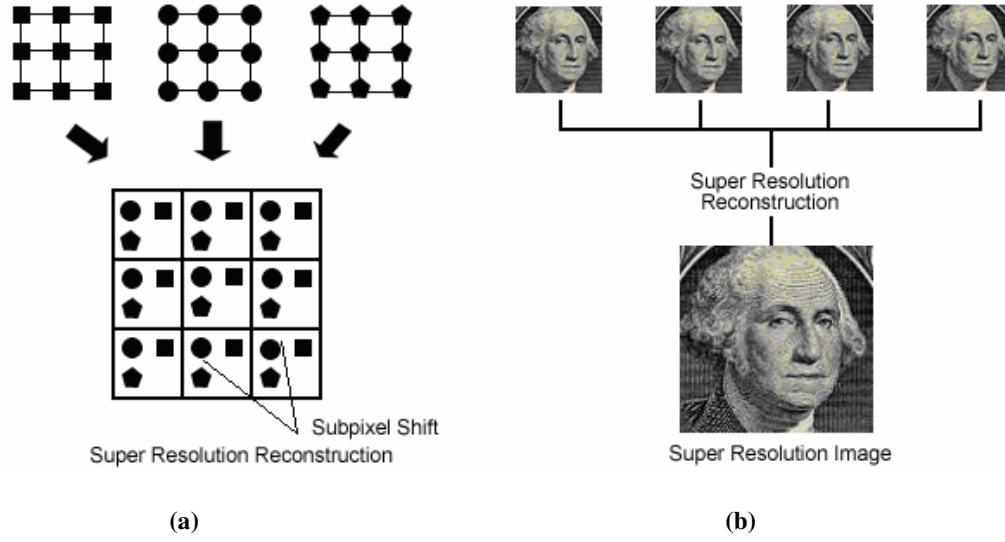
## **1.2 Problem Statement**

Due to many challenges such as pose difference, illumination and expression changes, occlusions, accessories, distance from the image acquisition device which brings about low resolution images, unconstrained face recognition is still an unsolved problem. Compared with

other challenges, face recognition from low resolution data video reduces the performance of the existing systems significantly.

The goal of this thesis is to develop a system that uses SR as an intermediate step for face recognition in low resolution video and to analyze the face recognition rate improvements. Multi-frame SR is a practical solution that may increase frame resolution, which could in turn improve face recognition rates. The approach of this thesis uses a set of mutually unregistered low resolution face frames captured from video to construct a new frame which is higher in resolution (e.g., see Figure 1). In practice, such a combination of information from multiple images is not trivial. There are two main problems that need to be solved in a super resolution algorithm. First, all the input images need to be correctly aligned with each other on a common grid. Next, an accurate, sharp image has to be reconstructed from the gathered information. If one of these two steps is not done well, the resulting image is not acceptable, and no gain in resolution could be obtained.

A higher-resolution image is defined as an image with more resolving power. This means that an image that is obtained by merely up-sampling and interpolating a low-resolution image does not have a higher resolution than its original. The resolving power of an image can be increased by adding high-frequency information typically based on knowledge about the specific image model.



**Fig 1:** (a) Illustration of super resolution reconstruction;  
(b) An example of super resolution reconstruction.

Various techniques have been proposed to obtain a single high resolution image from many low resolution images to enhance the face recognition performance [2, 3, 4, 5, 6]. These pixel domain techniques can be divided into two main categories as follows: methods using face specific constraints [3,4,5] and methods without any face specific constraints [7]. Even though these SR techniques has been used for arbitrary image enhancements, usage of these in a video sequences has not been adequately explored. Frequency domain method by Patrick Vandewalle, et al. [8] states higher performance in SR image generation while spatial domain method by D. Keren et al. [9] claims average performance while achieving better speed performance. In this thesis, the frequency domain method by Patrick Vandewalle, et al. [8] and the spatial domain method by D. Keren et al. [9] will be examined for applicability in series of video frames to improve the face recognition performance. Additionally, performance of two SR implementations will be compared for facial recognition on image and video databases.

The system employs an implementation of principal components analysis based on Eigenfaces [10]. It is common practice to develop a principal components algorithm (PCA) to determine if a region of the image is, in fact, a face. PCA, which incorporates configuration



information of face into a reduced dimensional subspace model, enables faster computation and robustness against noise and motion estimation errors [11]. In order to recognize faces, detection of the face plays a major role and “Haar-like features” will be utilized in this system. The basic idea of this method is to first train a classifier with a number of sample views of the object (face in this context). Once the classifier is built, then it can be used to search for the desired object in an image or live video frame from a camera.

### **1.3 Outline**

The implementation of super resolution as a middle step to improve face recognition in low resolution video sequences is presented in this thesis (see Fig. 2). The system under development in this thesis will also investigate and compare the rate of face recognition improvement using frequency and spatial domain super resolution techniques.

The following chapter discusses prior research focusing on the multitude of techniques in super resolution. The development and analysis of face detection, super resolution algorithms and face recognition are detailed in Chapter 3. The details on the structure implementation and final tweaking are detailed in Chapter 4. The results obtained are examined in Chapter 5, followed with concluding discussion and suggestions for future work will be discussed in the final chapter.

## Chapter 2 Background

Takeo Kanade, is considered to have designed and implemented the first face recognition system in 1973, which used two computers to examine approximately 800 faces [12]. His goal was to examine and extract the eyes, nose, and mouth of the subjects. His research into examining faces excited interest in facial image processing, leading to the development of facial recognition as it is known today.

Serious activity in face recognition began in the late 80's and early 90's. One of the major reasons for this was that processing power was finally reaching the point where large amounts of image processing could be done without significant waiting time. Early image processing was all done on static images stored on a hard drive. It was not possible to integrate the algorithms into real time because the system simply couldn't do it fast enough. As the amount of processing power increases, so does the ability to run more and more complex algorithms.

Face recognition consists of a face detection and a face recognition part. In the detection part, faces are detected in the scene and their region of interest (ROI) is forwarded to the face recognition process where the found faces are matched to a database in order to recognize and identify them. This thesis introduces a middle step, which is a SR technique to improve the resolution of the ROI. Hence higher recognition rates could be achieved using low resolution inputs.

## 2.1 Face Detection & Recognition

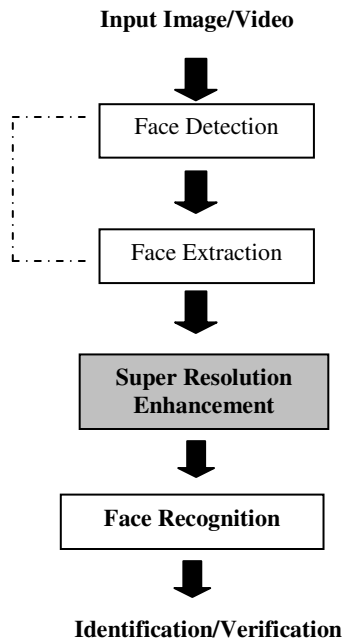
One of the most frequently researched tasks within the field of computer vision is that of human face detection and recognition. The human brain performs both tasks effortlessly and simultaneously, so even though it may seem that they are one and the same, detection and recognition are actually separate objectives requiring different approaches. Object detection is simply an act of determining whether or not an object is present; given an input image the output of an object detector will be a binary true/false answer. If an object is detected, the related task of localization the object within the image can often be performed at low cost by inference on the intermediate data utilized by the detector. Recognition on the other hand build upon an object detector by determining which object is present after separating the category into subclasses.

Face detection is a bottleneck for any face recognition system where the target is not held in a controlled state for scanning (crowd surveillance for example as opposed to in front of an ATM). Viola and Jones [14] proposed an efficient approach based on a combination of many weak filters to quickly capture possible faces in the image. A more modern version of this method was used in this research (OpenCV's [13] Haar Face Detector Lienhart [15]) for localizing ROI. In this thesis, this technique will be utilized for the face detection.

There are two main approaches for face recognition: holistic and geometric [16]. Geometric approaches dominated in the 1980's where simple measurements such as distance between the eyes and shapes of line connecting facial features [16] were used to recognize faces, while holistic methods became popular in the 1990's with the well known approach of Eigenfaces [17]. Even though holistic methods such as neural networks [19] are more complex to implement than their geometric counterparts, their application is much more straight forward, whereby an entire image segment can be reduced to a few key values for

comparison with other stored key values and no exact measures or knowledge, such as eye locations or the presence of facial hair (e.g. moustaches), needs to be known. The problem with this “grab all” approach is that noise, occlusions, such as glasses and any other non-face image attributes, could be learned by the holistic algorithm and become part of the recognition result, even though such factors are not unique to faces.

Although many face recognition algorithms have already been very successful with small databases [16,18,19], they were not aiming to solve small database recognition problem. Laurenz Wiskott *et al.* [7] present a system for recognizing human faces from signal images out of a large database containing one image per person. Faces are represented by labeled graphs, based on a Gabor wavelet transform.



**Fig.2:** General configuration of a face recognition system

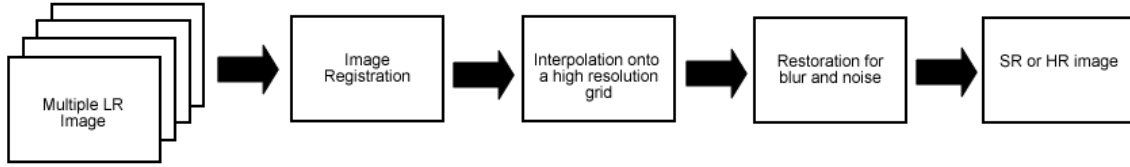
## 2.2 Super-resolution imaging

In most imaging applications, there is a big demand for high resolution (HR) images because the pixel densities within them are high. Therefore, HR images can not only give the viewer more pleasing pictures but also offer more details that maybe critical in different kinds of applications. The direct solution for obtaining HR images are mainly related to sensor manufacturing technique that attempts to increase the number of pixels per unit area by reducing the pixel size. However, the high cost high-precision optics and sensor may be a concern in general purpose commercial application. In addition, there exists a limitation to the reduction of pixel size due to the shot noise problem, which degrades the image quality severely [21].

In recent years, a resolution enhancement approach using signal processing techniques has been an active research area. As shown in Figure 1, SR image reconstruction algorithms investigate the relative sub-pixel motion information between multiple LR images and increase the spatial resolution by fusing them into a single frame. Therefore, inherent resolution limitation of the LR imaging system could be overcome. The major advance is cost effectiveness and utilizing existing LR imaging systems. As mentioned earlier, SR image reconstruction proved to be useful in many practical applications where multiple images of the same scene can be obtained, such as medical imaging, satellite imaging and video applications. Synthetic zooming of a ROI is another important application in surveillance, forensic, scientific, medical, and satellite imaging.

The idea of super resolution was first introduced in 1984 by Tsai and Huang [21] for multi-frame image restoration of band limited signals. A good overview of existing algorithms is given by Park et al. [22] and Borman and Settevenson [23]. Most SR methods proposed in the literature consist of three steps illustrated in Figure 3: image registration,

interpolation, and restoration. These steps can be implemented separately or simultaneously according to the reconstruction methods adopted.



**Fig 3:** Three basic steps of super resolution image reconstruction

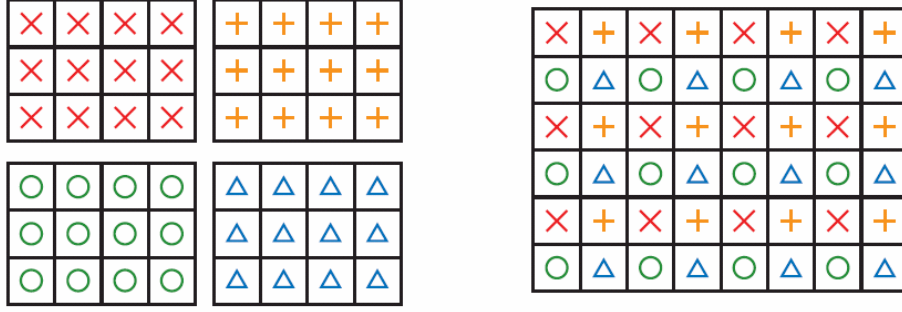
There are many different methods for SR reconstruction [23],[24], and these methods can divide into three different categories such as frequency domain method, spatial domain method and learning-based methods. In this thesis, derivatives of the first two methods will be used and their performance in the context of face recognition will be compared. Frequency domain method uses the frequency aliasing that is introduced during the acquisition of the LR images. Tsai et al. first proposed this type of method in [21], and was later extended in [25],[26]. Most of the frequency domain registration methods are based on the fact that two shifted images differ in frequency domain by a phase shift only, which can be found from their correlation.

Spatial domain methods improve the SR formulations by incorporating spatially varying noise and blurring processes as well as complex motion relationships between the LR images. Irani et al. [28] suggested iterative backward projection for SR, which projects the errors between the simulated and observed LR images back into the HR images until the errors reach a minimum. In [29], Schultz et al. proposed a maximum a posteriori framework for SR reconstruction that uses an edge preserving Huber-Markov random field for the image prior. This framework was then extended in [30] to simultaneously restore, register and interpolate the LR images by using cyclic coordinate-descent minimizing procedure. Karen *et al.* [9] developed an iterative planar motion estimation algorithm based on Taylor expansions. A pyramidal scheme was used to increase the precision for large motion parameters; Bergaen

*et al.* developed a hierarchical framework to estimate motion in a multi-resolution data structure [31].

If we want to increase the resolution of an image using super-resolution techniques, we essentially need to be able to distinguish more details in the final image. By adding images of the same scene, we try to add information to the reproduction. Typically this information is high frequency content of the scene.

There are different ways to add such high frequency information to an image. If we know that the image is of a certain type (faces, text, drawings, etc.), we can use that knowledge to add frequency content. Such an approach is called a model-based approach. For example, if we know that the images represent printed text, we can try to recognize characters, and replace them by sharp, high quality characters. The knowledge of the image model allows us to compute high frequency information. In this thesis, we will investigate abstract approaches to super-resolution. They use other information than a precise image model, and are therefore applicable to more general types of images. More specifically, we will compute the high frequency information from the aliasing that is present in the images. Super-resolution techniques use a number of low resolution input images to generate a high resolution image. This assumes that there are some (small) differences between the input images. Most often, these differences are caused by small camera movements. In an ideal situation, we could assume that of four images taken, the second to fourth image have a horizontal, vertical, and diagonal shift of a half a pixel compared to the first image. As shown in Figure 4, the pixels from the first image can then be interleaved with pixels from the three other images, and double the resolution of the image (in both dimensions) . In general, however, the shifts between the images are not exactly half a pixel and can be any arbitrary value, but the scope of this thesis address face recognition which will support only sub pixel level movement.



**Fig 4:** Ideal super-resolution setup. Four images are taken with relative shifts of half a pixel in horizontal, vertical, and diagonal directions (left). Their pixels can then be interleaved to generate a double resolution image (right).

Additionally, there has been research effort in using super resolution enhancement of images for face recognition. Two of the significant investigations were done by Lin, Cook, Chandran and Sridharan [32] and Sezer, Altunbasak and Ercil [33]. Lin, Cook, Chandran and Sridharan. introduced a novel optical flow based super resolution face recognition system. Sezer, Altunbasak and Ercil introduced super resolution in feature domain, or face subspace, instead of pixel domain claiming advantages in robustness against noise and motion estimation errors. Optical flow based super resolution method limits the lowest resolution that could be used because of the PCA based face recognition used. As an example, in order to see performance impact on face recognition, extreme LR images should be super resolved.

## 2.3 Application Domains

Every day, we are facing new products of technology, prompting us to enter our PIN code or passwords for financial transactions on the internet or to get cash from ATM, even to use our cell phone SIM card, or a dozen of others, e.g. to access internet and so on.



Therefore, the need for reliable methods of biometric personal identification is obvious. In fact, there are such reliable methods like fingerprint analysis, retinal or iris scans, however these methods rely on the cooperation of the participant. Face recognition systems, on the other hand, can perform person identification without the cooperation or knowledge of participant which is advantageous in some applications, such as surveillance, suspect tracking and investigation. Typical application of face recognition systems can be listed in at least four main categories [21]:

- (i) Entertainment: Video Game, Virtual Reality, Training Programs, Human-Robot-Interaction, Human-Computer-Interaction.
- (ii) Smart Cards: Driver's Licenses, Entitlement Programs, Immigration, National ID, Passports, Voter Registration, Welfare Fraud.
- (iii) Information Security: TV Parent Control, Personal Device Logon, Desktop Logon, Application Security, Database Security, File Encryption, Intranet Security, Internet Access, Medical Records, Secure trading Terminals.
- (iv) Law Enforcement and Surveillance: Advance Video Surveillance, CCTV Control, Portal Control, Post-Event Analysis, Shoplifting, Suspect Tracking and Investigation.

These commercial and law enforcement applications of face recognition systems vary according to the format of the imagery ranging from static images and controlled-format face images to uncontrolled video sequences. In law enforcement and surveillance applications resolution is extremely vital when using for different object recognition.

Super-resolution methods can be used to create high resolution still pictures or video from low resolution video sequences. In surveillance cameras, additional details can be revealed by combining multiple video frames to create a single high resolution image. The same techniques can be applied to improve the resolution of existing (low resolution) video content for use in high definition television sets. While existing research has addressed

several issues in the analysis of video, little work has been done in the area of better information acquisition of high resolution face images. Resolution of images or frames is a key factor in face recognition since it is one of the limiting parameters in digital camera design. Super resolution (SR) is one of the key techniques exercised to improve low resolution images and could be used in low resolution (LR) surveillance video sequences to improve the face recognition rates. Even though SR has been used extensively in many applications to improve the quality of images usage in the video sequences has been minimal. The methods described in the following chapters can therefore also be applied to such problems.

## Chapter 3    Theoretical Development

In order to achieve the objective of this thesis, which is to improve face recognition rates using super resolution techniques in video sequences, it is obvious that face detection and recognition are needed steps. Face detection is done using OpenCVs' Haar face detector and Eigenfaces with PCA algorithm was utilized for face recognition. The algorithm was implemented and tested using three types of databases namely image, video and web camera captured video in order to obtain efficient, yet accurate results.

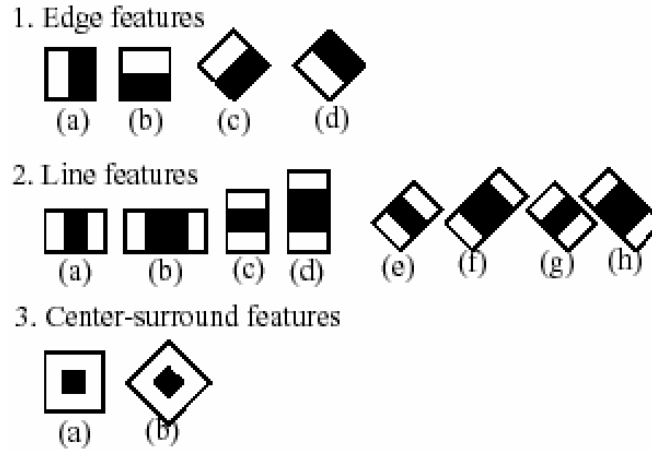
### 3.1 Face Detection using Haar-Like Features

This is a widely used algorithm with a strong mathematical base. Paul Viola was the first to develop the Haar cascade object detector [14] that was improved by Rainer Lienhart [15]. The idea here is to first train a classifier with a number of sample views of object. This approach to detecting objects (in this thesis faces) in image combines four key concepts:

- Simple rectangular features, called Haar Features.
- An integral image for rapid feature detection.
- The AdaBoost machine learning method.
- A cascaded classifier to combine many features efficiently.

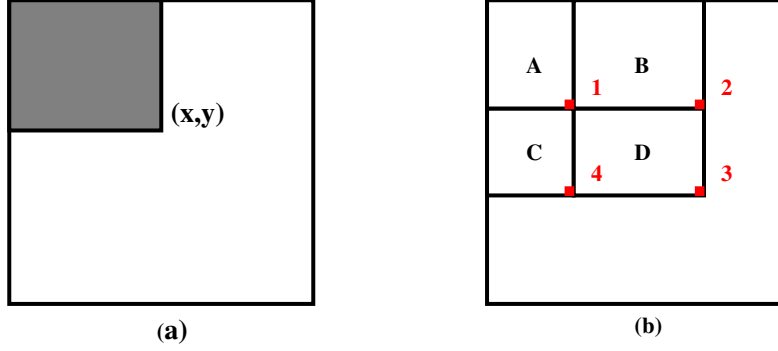
The features that Viola and Jones used are based on Haar wavelets which are single-wavelength square waves (one high interval and one low interval). In two dimensions, a square wave is a pair of adjacent rectangles one light and one dark. The actual rectangle combinations used for visual object detection are not true Haar wavelets. Instead, they contain rectangle combinations better suited to visual recognition tasks. Because of that

difference, these features are called Haar features or Haar like features, rather than Haar wavelets. Figure 5 shows the features that OpenCV uses [13].



**Fig 5:** Some Haar-like features from [13]

The presence of a Haar feature is determined by subtracting the average dark-region pixel value from the average light-region pixel value. If the difference is above a threshold (set during learning), that feature is said to be present. To efficiently determine the presence or absence of hundreds of Haar features at every image location and at several scales Viola and Jones used a technique called an integral image. Generally, "integrating" means adding small units together, but in this case the small units are pixel values. The integral value for each pixel is the sum of all the pixels above it and to its left. Starting at the top left and traversing to the right and down, the entire image can be integrated with a few integer operations per pixel. As Figure 6 shows, after integration, the value at each pixel location,  $(x,y)$  contains the sum of all pixel values within a rectangular region that has one corner at the top left of the image and the other at location  $(x,y)$ . The average pixel value in this rectangle is found by dividing the value at  $(x,y)$  by the rectangles area.



**Fig 6:** The integral image trick. After integrating, the pixel at (x,y) contains the sum of all pixel values in the shaded rectangle. The sum of pixel values in rectangle D is  
 $(x_4, y_4) - (x_2, y_2) - (x_3, y_3) + (x_1, y_1)$ .

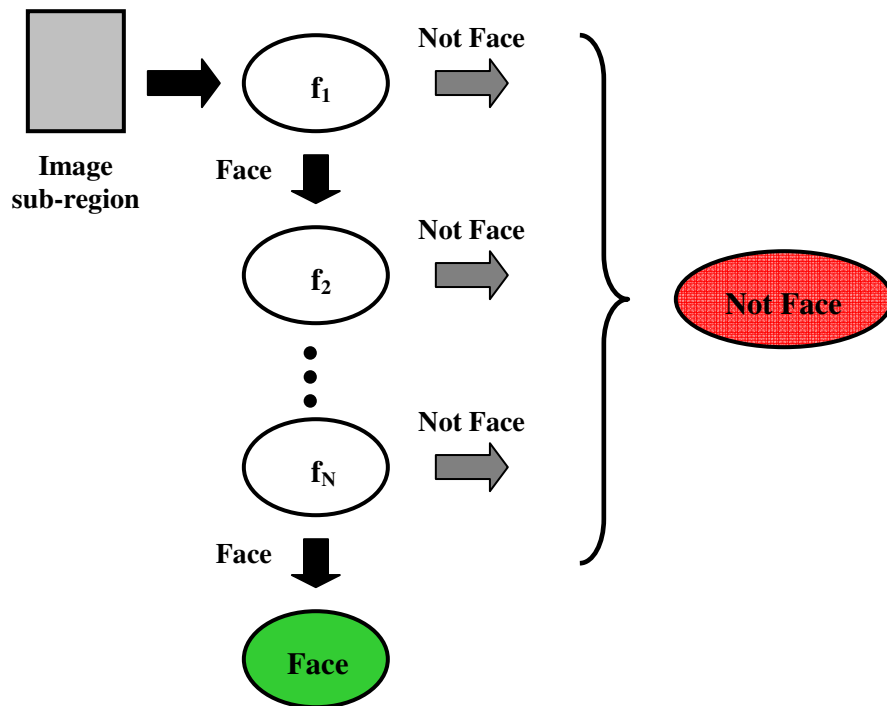
For a rectangle which is not located at the upper left corner of the image, as in Figure 6(b), the sum of pixel values in D can be calculated as the combined rectangle, A+B +C+D, minus the sums in rectangles A +B and A+C, plus the sum of pixel values in A. In other words,

$$D = A + B + C + D - (A+B) - (A+C) + A$$

Conveniently, A+B+C+D is the integral image value at location 4, A+B is the value at location 2, A+C is the value at location 3, and A is the value at location 1. So, with an integral image, the sum of pixel values for any rectangle in the original image is obtained with just three integer operations:

$$(x_4, y_4) - (x_2, y_2) - (x_3, y_3) + (x_1, y_1)$$

To select the specific Haar features to use and to set threshold levels, Viola and Jones use a machine-learning method called AdaBoost [13]. AdaBoost combines many "weak" classifiers to create one "strong" classifier. "Weak" here means the classifier only gets the right answer a little more often than a random guessing, which is not acceptable. If all these are weak classifiers and each one "pushed" the final answer a little bit in the right direction, a strong, combined force for arriving at the correct solution could be achieved. As in Figure 7, AdaBoost selects a set of weak classifiers consisting of one Haar feature to combine and assigns a weight to each and this weighted combination is the strong classifier.

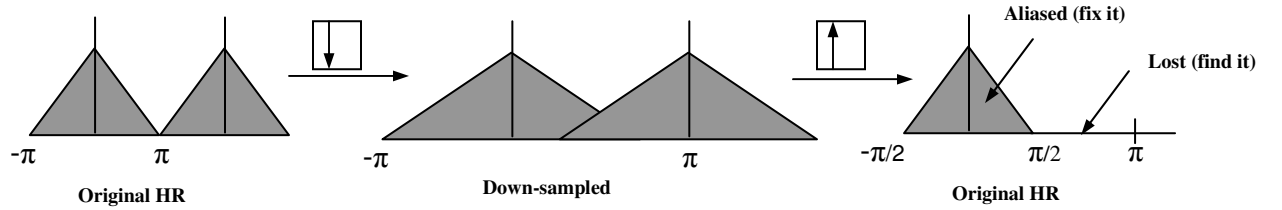


**Fig 7:** The classifier cascade is a chain of single-feature filters. Image sub-regions that make it through the entire cascade are classified as 'Face.' All others are classified as "Not Face."

The threshold for each filter is set low enough that it passes all, or nearly all, face examples in the training set, where the training set is a large database of faces, maybe a thousand or so. During use, if any one of these filters fails to pass an image region, that region is immediately classified as "Not Face". When a filter passes an image region, it goes to the next filter in the chain. Image regions that pass through all filters in the chain are classified as "Face". Viola and Jones dubbed this filtering chain a cascade. The order of filters in the cascade is determined by weights that AdaBoost assigns. The more heavily weighted filters used first in the cascade chain to eliminate non-face image regions as quickly as possible.

## 3.2 Super Resolution

Before delving into solutions, let's first try to better understand our objective. Let us take a high-resolution image and down-sample it (without low-pass filtering). Hence, there will be aliasing in the down sampled images. In real-life these down-sampled images will be given with aliasing and our goal is to create an HR image from them. As shown in the Figure 8 below, it boils down to finding the lost high-freq components and fixing the aliased low-freq values in the HR image.



**Fig 8:** Graphical representation of super resolution problem.

As mentioned in Chapter 2, there are many techniques addressing solutions for the problem. The super resolution algorithms described below were chosen for their superior results and reputability. One algorithm utilizes the frequency domain approach while the other exploits spatial domain. The approach of Patrick Vandewalle, *et al.* [8] was used in the thesis as the frequency domain method and the method by D. Keren *et al.* [9] was used as the spatial domain approach. The theoretical development of each technique is discussed next.

### 3.2.1 Frequency Domain Approach

Fourier based image registration methods only allow global motion in a plane parallel to the image plane. Therefore, motion between two images can be described as a function of three parameters that are all continuous variables: horizontal and vertical shifts  $x_{1,h}$  and  $x_{1,v}$

and a planar rotation angle  $\theta_1$ . The frequency domain approach allows us to estimate the horizontal and vertical shift and the (planar) rotation separately. Hence, a continuous, two-dimensional reference signal  $f_0(x)$  and its shifted and rotated version  $f_1(x)$  could be represented as:

$$f_1(x) = f_0(R(x + x_1)),$$

$$\text{where } x = \begin{pmatrix} x_h \\ x_v \end{pmatrix}, x_1 = \begin{pmatrix} x_{1,h} \\ x_{1,v} \end{pmatrix}, R = \begin{pmatrix} \cos \theta_1 & -\sin \theta_1 \\ \sin \theta_1 & \cos \theta_1 \end{pmatrix}$$

This can be expressed in Fourier domain as,

$$\begin{aligned} F_1(u) &= \iint_x f_1(x) e^{-j2\pi u^T x} dx \\ &= \iint_x f_0(R(x + x_1)) e^{-j2\pi u^T x} dx \\ &= e^{-j2\pi u^T x_1} \iint_{x'} f_0(R(x')) e^{-j2\pi u^T x'} dx', \end{aligned}$$

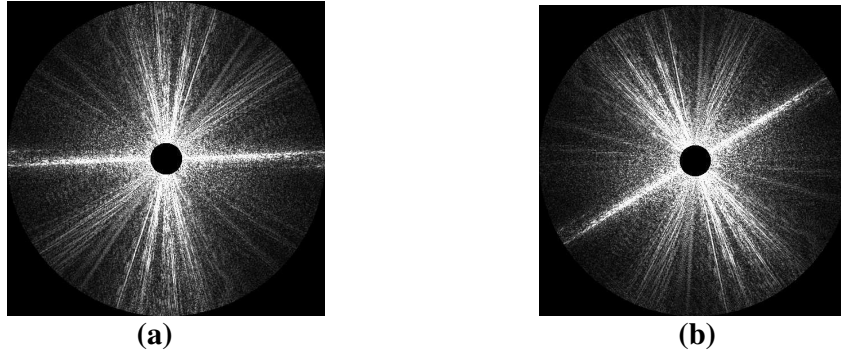
where  $F_1(u)$  is the two-dimensional Fourier transform of  $f_1(x)$  and the coordinate transformation  $x' = x + x_1$ . After another transformation  $x'' = Rx'$ , the relation between the amplitudes of the Fourier transforms can be computed as

$$\begin{aligned} F_1(u) &= \left| e^{-j2\pi u^T x_1} \iint_x f_1(x) e^{-j2\pi u^T x'} dx \right| \\ &= \left| \iint_{x'} f_1(x) e^{-j2\pi u^T x'} dx \right| \\ &= \left| \iint_{x''} f_0(x'') e^{-j2\pi u^T (R^T x'')} dx'' \right| \end{aligned}$$



$$= |F_0(Ru)|$$

The above equation illustrates that  $|F_I(u)|$  is a rotated version of  $|F_O(u)|$  over the same angle  $\theta_I$  as the spatial domain rotation (see Figure 9). The magnitudes  $|F_O(u)|$  and  $|F_I(u)|$  do not depend on the shift values  $x_I$ , because the spatial domain shifts only affect the phase values of the Fourier transforms. Therefore, rotation angle  $\theta_I$  can be estimated from the amplitudes of the Fourier transforms  $|F_O(u)|$  and  $|F_I(u)|$ . After compensation for the rotation, the shift  $x_I$  can be computed from the phase difference between  $F_O(u)$  and  $F_I(u)$ .



**Fig 9:** The amplitude of the Fourier transform of an image is rotated over the same angle ( $\theta_I = 25^\circ$ ) as the spatial domain image. (a) Fourier transform amplitude of the original image. (b) Fourier transform amplitude of the rotated image.

### 3.2.1.1 Rotation estimation

If  $|F_O(u)|$  and  $|F_I(u)|$  are transformed in polar coordinates, the rotation over the angle  $\theta_I$  is reduced to a (circular) shift. The angle  $\theta_I$  could be calculated as the phase shift between Fourier transform of the polar spectra  $|F_O(u)|$  and  $|F_I(u)|$ , as it was done by Marcel et al and Reddy and Chatterji [34]. This requires a transformation of the spectrum to polar coordinates. Mainly for the low frequencies, which generally contain most of the energy, the interpolations are based on very few function values and thus introduce large approximation errors.

### 3.2.1.2 Shift estimation

A shift of the image parallel to the image plane can be expressed in Fourier domain as a linear phase shift:

$$\begin{aligned} F_1(u) &= \iint_x f_1(x) e^{-j2\pi u^T x} dx = \iint_x f_0(x + x_1) e^{-j2\pi u^T x} dx \\ &= e^{-j2\pi u^T x_1} \iint_x f_0(x) e^{-j2\pi u^T x} dx = e^{-j2\pi u^T x_1} F_0(u) \end{aligned}$$

It is well known that the shift parameters  $x_1$  can thus be computed as the slope of the phase difference  $\angle (F_1(u)/F_0(u))$  [34]. Phase differences using a least squares method has been used to make the solution less sensitive to noise.

### 3.2.1.3 Reconstruction

When the low resolution images are accurately registered, the samples of the different images can be combined to reconstruct a high resolution image. In the reconstruction algorithm, the samples of the different low resolution images are first expressed in the coordinate frame of the reference image. Then, values will be interpolated on a uniform high resolution grid. Bi-cubic interpolation was used because of its low computational complexity and good results.

An important question is how many images are required to increase the resolution by a certain factor, The optimal number of low resolution images to use for a super-resolution algorithm depends on many parameters: registration accuracy, imaging model, total frequency content, etc. Intuitively, there is a trade-off between two effects. On one hand, the more images there are, better the reconstruction should be. On the other hand, there is a limit to the improvements that can be obtained: even from a very large number of very low

resolution images of a scene, it will not be possible to reconstruct a sharp, high resolution image. Blur, noise, and inaccuracies in the signal model limit the increase in resolving power that can be obtained. In this algorithm, motion estimation is limited to sub sampling by a factor less than two in both dimensions. It cannot be increased more, because higher frequencies are not present in the measured images. When an image resolution is increased by a factor two in both dimensions, it results in an image with four times the original number of pixels. In frequency domain, we could say that in order to compute all frequency components of the new image, four times as many measurements would be needed. This completes the overview of frequency domain super-resolution algorithm. The spatial domain approach is discussed next.

### ***3.2.2 Spatial Domain Approach***

In this section we consider a simple, computationally efficient approach to constructing super-resolution images from an image sequence based on spatial domain interpolation. Though this approach may initially appear attractive, it is overly simplistic as it does not take into consideration the fact that samples of the low resolution images do not result from ideal sampling, but are, in fact, spatial averages. The result is that the reconstructed image does not contain the full range of frequency content that can be reconstructed given the available low-resolution observation data. Keren, Peleg and Brada [9] describe a spatial domain approach to image registration using a global translation and rotation model, as well as a two stage approach to super-resolution reconstruction. The motion estimation techniques and the first stage of the super resolution reconstruction method, which is a simple interpolation technique, are as follows.

Consider the continuous image pair  $f_a(x_1, x_2)$  and  $f_b(x_1, x_2)$  related by global translation and rotation as

$$f_b(x_1, x_2) = f_a(x_1 \cos \theta - x_2 \sin \theta + \Delta_{x_1}, x_2 \cos \theta + x_1 \sin \theta + \Delta_{x_2})$$

Where  $\Delta_{x_1}$ ,  $\Delta_{x_2}$ ,  $\theta$  are the translation parameters in the  $x_1$  dimension, translation in the  $x_2$  dimension and rotation, respectively. By expanding the trigonometric functions in a second order Taylor series and then expanding  $f_a$  as a Taylor series about  $(x_1, x_2)$ , an approximate error  $\epsilon(\Delta_{x_1}, \Delta_{x_2}, \theta)$ , between  $f_a$  and  $f_b$  may be determined as,

$$\epsilon(\Delta_{x_1}, \Delta_{x_2}, \theta) = \sum \left[ f(x_1, x_2) + \left( \Delta_{x_1} - x_2 \theta - x_1 \frac{\theta^2}{2} \right) \frac{\partial f}{\partial x_1} + \left( \Delta_{x_2} + x_1 \theta - x_2 \frac{\theta^2}{2} \right) \frac{\partial f}{\partial x_2} \right]^2$$

The summation is over overlapping portions of  $f_a$  and  $f_b$ . The minimum of  $\epsilon(\Delta_{x_1}, \Delta_{x_2}, \theta)$ , is found by differentiation and equating to zero, yielding a set of three equations in three unknowns. This system is solved using an iterative method, utilizing a Gaussian pyramid to expedite processing. The result is the registration parameters  $\Delta_{x_1}$ ,  $\Delta_{x_2}$ ,  $\theta$ . It is reported that the approach yields parameters with sub-pixel accuracy provided is small, due to the small angle approximation to the trigonometric functions describing the image rotation.

Given  $R$  observed frames the registration procedure described above is used to extract  $\Delta_{x_1}$ ,  $\Delta_{x_2}$  and  $\theta$  for each of  $R-1$  images relative to a chosen reference image. The images are registered and a high resolution reconstruction grid is imposed on the “stack” of observed images. Each pixel in the high resolution image is chosen as an outlier trimmed mean of the value of the set of observation image pixels the centers of which fall within the area of the high resolution pixel under consideration. Post-processing of the interpolated image may be effected using a de-blurring filter, under the assumption of a Gaussian blur.

The resulting image, which is a composite of the observed low resolution images, does not significantly extend the frequency domain content of the reconstructed image beyond that of any individual low-resolution image, since the reconstruction is predominantly the result of linear combination of the data, which is known not to extend the frequency content of the signal. Though the paper does contribute early ideas concerning image registration, the image interpolation procedure is overly simplistic in its approach to the super-resolution problem. No attempt is made to de-alias the input data. In effect, only a fraction of the available *information* is utilized in the reconstruction.

### **3.3 Eigenface-based facial recognition**

The information theory approach of encoding and decoding facial images can be summarized as extracting the relevant information in a face image, encoding it as efficiently as possible and comparing it with database of similarly encoded faces. The encoding is done using features which may be different or independent from the distinctly perceived features like eyes, ears, nose, lips, and hair. Eigenfaces are such an encoding mechanism for face images [10].

One of the major advantages of Eigenfaces-based recognition is the ease of implementation. Furthermore, no knowledge of geometry or specific feature of the face is required; and only a small amount of work is needed in preprocessing for any type of face images. However, a few limitations are demonstrated as well. It is applicable only to front views. Second, as it is addressed in [35] and many other face recognition related literatures, it demonstrates good performance only under controlled background, and may fail in natural scenes. These disadvantages are outside the scope of this thesis which address only on frontal poses in a controlled environment.

Independent Component Analysis (ICA) was initially developed to separate individual signals from a combination [36]. It breaks a complicated signal into additive subcomponents which are statistically independent. It was first used for face recognition by Bartlett and Sejnowski [37]. ICA differs from PCA in that it uses higher order moments in order to separate the data. PCA uses 2<sup>nd</sup> order moments and uncorrelated the data. When used for face recognition, ICA assumes that a face is a combination of a set of unknown source images. These independent images can be used to identify a particular subject. While ICA has been shown to be more effective at distinguishing differences between images of a set, it does not generalize well for images which are not in the training set.

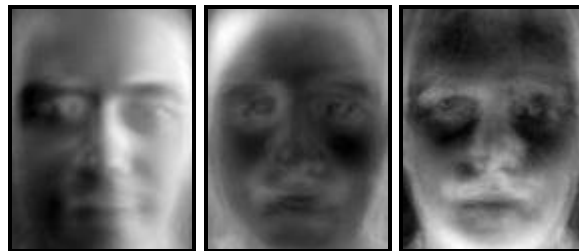
Every human face has a unique, yet similar topological structure. A method which examines this structure in order to determine the identity of the face is Elastic Bunch Graph Matching (EBGM) [38]. In this system, faces are represented as graphs, which have nodes at important features such as the eyes, nose, and mouth. The edges between the nodes are weighted with the distance between the nodes. In addition, each node contains information about its feature in the form of Gabor wavelets. While this method differs greatly from the other methods examined above, it does not achieve the same levels of performance.

Bayesian statistics have been used in conjunction with both Eigenfaces and Fisherfaces to improve the performance of the feature vectors in face classification. This approach uses a statistical classifier to help determine the importance of the components of the feature vector, allowing for more accurate face classification [41, 42]. However, the use of Bayesian statistics is heavily dependant on the underlying feature set selection and can only marginally improve performance. Thus, Bayesian statistics will not be employed in this thesis.

One of the most promising methods explored in face recognition arena is called Linear Discriminant Analysis (LDA). LDA reduces the dimensionality of a face image in a

similar method to PCA, yet instead of producing orthogonal vectors, it produces vectors which linearly separate the data as much as possible [43]. Therefore, it theoretically produces a set of vectors which are more effective at face classification than Eigenfaces. The LDA vectors are similar to Eigenfaces and are often called Fisherfaces since they were developed by Ronald Fisher. Unfortunately, in order to train such a system to be more effective than PCA, all of the variables which will be encountered by test subjects must be in the training set. Such variables include lighting, pose angle, and expression. Since the system will be run in a dynamic environment, ensuring incorporation of all of the variables may be impossible.

Mathematically, principal component analysis approach treats every image of the training set as a vector in a high dimensional eigenspace. The eigenvectors of the covariance matrix would efficiently incorporate the variation amongst the face images. Now each image in the training set would have its contribution to the construction of the eigenvectors called “Eigenfaces”. When saving or viewing an Eigenface through the system, a linear scaling is done whereby the lowest value is scaled to 0 and the highest to 255. These Eigenfaces look like ghostly images and some of them are shown in Figure 10. In each Eigenface some sort of facial variation can be seen which deviates from the original image.

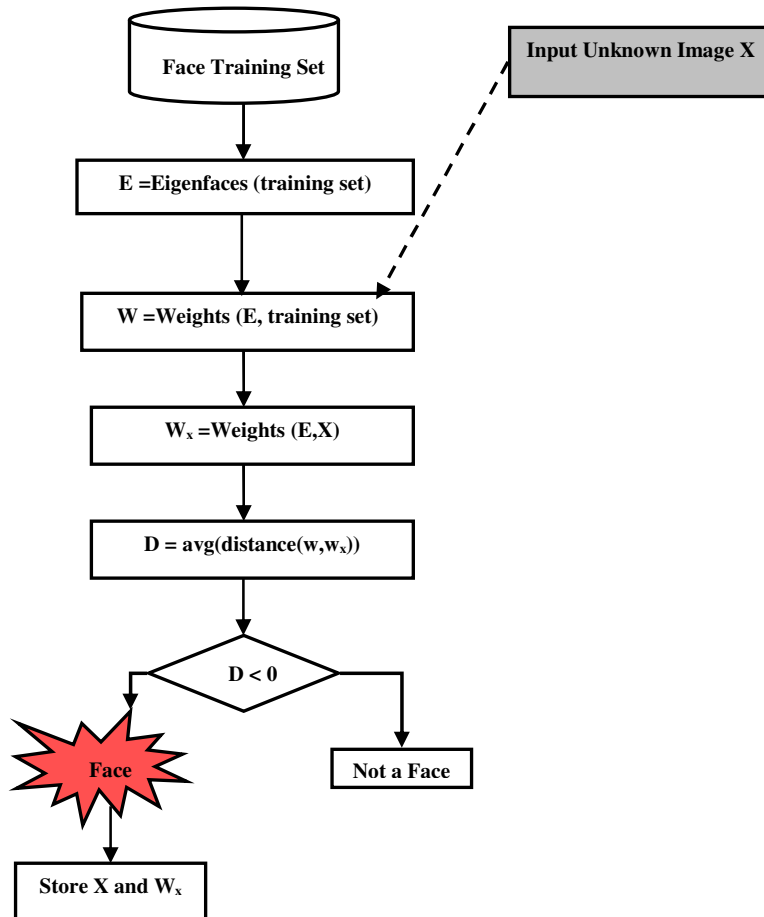


**Fig 10:** Three Eigenfaces

The high dimensional space with all the Eigenfaces is called the image space (feature space). Also, each image is actually a linear combination of the Eigenfaces. The amount of

overall contribution of one Eigenface is represented by the Eigenvalue associated with the corresponding eigenvector. If the Eigenfaces with small Eigenvalues are neglected, then an image can be a linear combination of reduced number of these Eigenfaces.

When the face image to be recognized (known or unknown), is projected on this face space, we get the weights associated with the Eigenfaces, that linearly approximate the face or can be used to reconstruct the face. Next these weights are compared with the weights of the known face images so that it can be recognized as a known face. In simpler words, the Euclidean distance between the image projection and known projections is calculated; the face image is then classified as one of the faces with minimum Euclidean distance. Figure 11 shows a high level overview of the Eigenface-based recognition.



**Fig 11:** High-level functions of the Eigenface-based facial recognition algorithm



### 3.4 Image and Video Databases

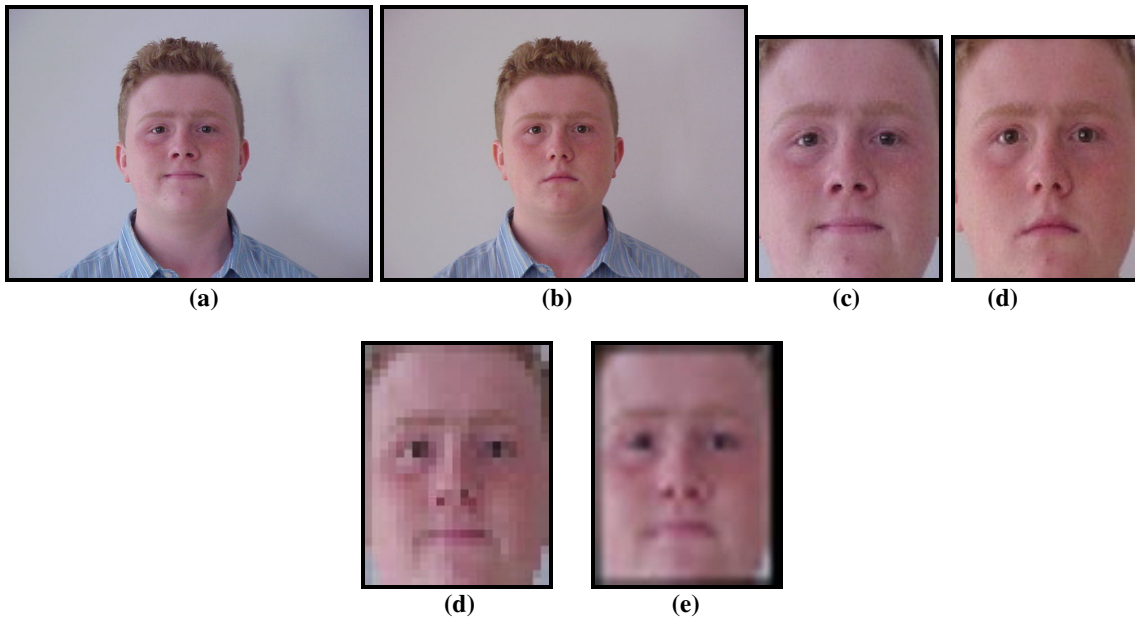
In order to obtain the sheer volume of images needed to develop, train, and test the system, the following databases were selected. Two image databases were used: one was the Indian face database [44] and the other was the CVL face database [45]. The Indian face database was developed in 2002 and maintained by the Indian Institute of Technology, Kanpur. This database contains images of 60 distinct subjects with eleven different poses for each individual. When available, a few additional images are also included for some individuals. All the images have a bright homogeneous background and the subjects are in an upright, frontal position. In order to develop a cohesive set of test and train images, two images of each subject from different situations were examined and selected. Since the objective of the testing would be to match an image of a subject to another image of the same subject, two images of each individual were needed. The two images were selected to simulate as if it was captured through close by video frames. Therefore the images are similar, yet not identical.

The second image database used was CVL face database [45] which contains 114 subjects with 7 different poses for each subject and all subjects are in an upright frontal position. All the images were taken using Sony Digital Mavica camera under uniform illumination, no flash, and with projection screen in the background. As for the Indian face database, two images with different poses were selected for testing and training.

The video face database used was VPA super resolution video face database [46] which consists of high resolution frontal face images of 32 people and 32 low resolution videos. Face videos contain just translational movement of each face shot by SONY DVR camera from a distance in ambient light and with a video length of 4 seconds. In Chapter 5 more information on preparation of databases will be discussed.

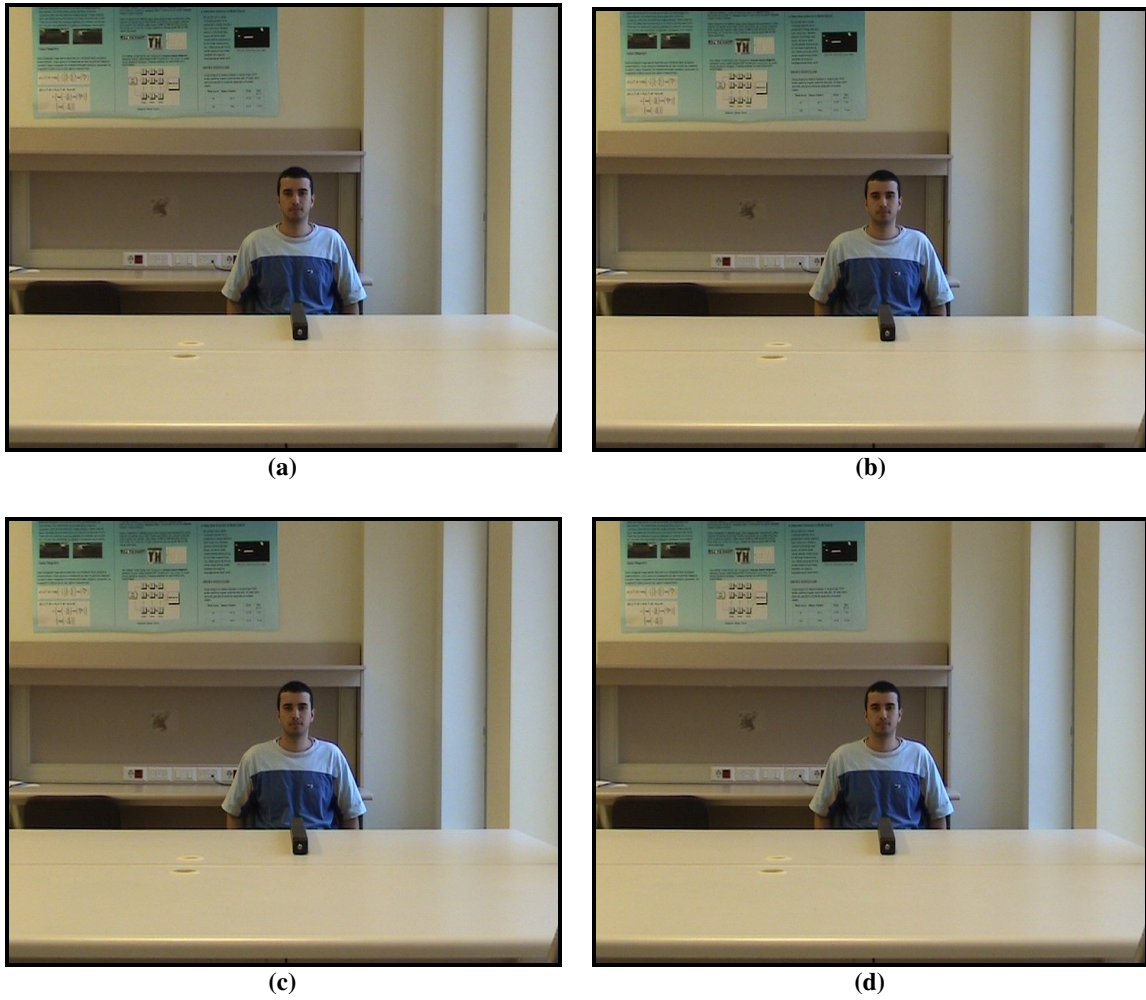
### 3.4.1 Data Preparation

For both image and video databases one training set was created and four databases were tested against these training data. Therefore one image was selected from each subject for training from all four databases which accounted for 198 training images. Testing images were hand picked from image databases. First, selected images from each subject were categorized as a HR image. Once the HR images were selected, face detection and auto cropping was run to create a set of cropped 90x120 HR images where faces were cropped at the top of the forehead, bottom of the chin, and at the base of the ears (ROI) and images were equalized. Figure 12(b) and 12(c) shows results of cropping. The same image was then used to generate four, 45x60 LR images and 16, 23x30 LR images. These steps were followed in both image databases to create two test datasets. Also, a randomly selected LR image was used to generate two images matching to the size of HR image, one using bicubic interpolation and another without interpolation by just zooming. Both of these images will be used in testing to compare against SR face recognition rates.



**Fig 12:** Original CVL database images and the resulting cropped images (a) selected original image for training (b) selected original image for testing (c) cropped version of training image (d) cropped version of test image (e) zoomed version of test image (f) bicubic interpolated version of test image

Each video is a 720x576 in size and 4 seconds in length which contained 100 frames. In each video, the subject (person) does a very slight movement of the head due to reading a text which could be easily assumed as a sub pixel level between frames. Figure 13 shows four consecutive frames from 15<sup>th</sup> to 19<sup>th</sup> frames of a selected subject which demonstrates that the amount of movement is at the sub pixel level.



**Fig 13:** Frame captures from four consecutive frames.  
(a) 15<sup>th</sup> frame (b) 16<sup>th</sup> frame (c) 17<sup>th</sup> frame (d) 18<sup>th</sup> frame

Testing frames for video databases were selected as follows. First, face detection and auto cropping was run on LR video to grab and crop four, 45x60 LR images and 16, 23x30 LR frames. Then the images were equalized. The cropped area was the ROI as in the image database. The respective image from the database was used as the HR frame and used the same cropping mechanism to create a 90x120 equalized image.

Images for the training set were hand picked to have a different pose from the originally selected image for HR test data set. Preprocessing steps of cropping and equalization were the same as for the test images. Video frames for the training set were selected from different frames in order to capture a different pose from the HR test data. Therefore, each subject on the HR test data set had a respective subject image in the training dataset with a different pose for both video and image databases.

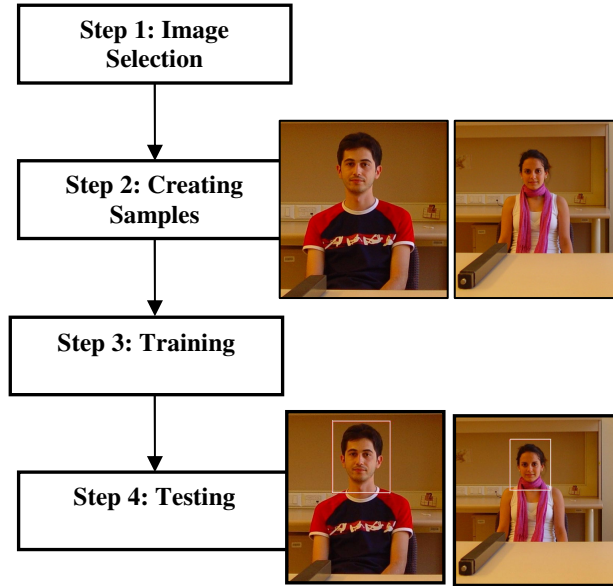
## Chapter 4 System Development

Developing an implementation of the face detection and recognition algorithms to work with the super resolution techniques is not a simple task. In order to obtain enhanced results in super resolution it is vital to have sub pixel level movements in video frames. In addition, to replicate low resolution video frames, low resolution images were generated using high resolution image databases. Further details into implementation of algorithms and tweaking are discussed below.

### 4.1 Face Detection

OpenCV is an open source computer vision library originally developed by Intel [47]. As discussed in Chapter 2, the “*Boosted Cascade of Simple Features*” object detection algorithm introduced by Viola and Jones was built into OpenCV and will be utilized in this thesis. Steps shown in Figure 14 were followed to train the Harr classifier. The first step is to select positive/negative samples out of the training image database. Negative samples are taken from arbitrary images to avoid the camera variation when testing is performed on different databases. These images do not contain face representations. Negative samples are passed through background description file where each line contains the filename (relative to the directory of the description file) of the negative sample image. This file was created once manually and was used across databases for face detection. When selecting positive samples, samples were selected from different reflections, illuminations and backgrounds of different people to mimic real world scenes. First and foremost, the face was localized to include eyes,

eyebrow, nose and mouth and find the numbers, locations, face width and height. The face was localized so that the rectangle is close to the object border.



**Fig 14:** Steps of the Harr training process are shown above

This collection is described by a text file similar to the background description file. Each line of this file corresponds to a collection image. The first element of the line is the image file name and followed by the number of object instances. The following numbers are the coordinates of bounding rectangles (x, y, width, height). An example of the description file is below:

```

img/img1.jpg 1 140 100 45 45
img/img2.jpg 2 100 200 50 50 50 30 25 25

```

The description file can be explained as image *img1.jpg* that contains a single object instance with bounding rectangle (140, 100, 45, 45) and image *img2.jpg* that contains two object instances. Once negative and positive images are marked, the “*opencv-createsamples*” utility can be used to create binary file for positive and negative samples. An example command line argument for running the utility is below:

```
./src/opencv-createsamples -info info.txt -vec infovec.vec -num 2400 -w 25  
-h 25
```

It will take few minutes to create a vec file(binary file) according to the number of samples. After creation of the vec file, training could be done which is the most important part for detection.

The next step is the training of the classifier which was performed using the “*opencv-haartraining*” utility.

```
./src/.libs/lt-opencv-haartraining -data /TrainingFolders/ -vec infovec.vec  
-bg bg.txt -npos 2200 -nneg 2200 -nstages 30 -mem 300 -w 25 -h 25
```

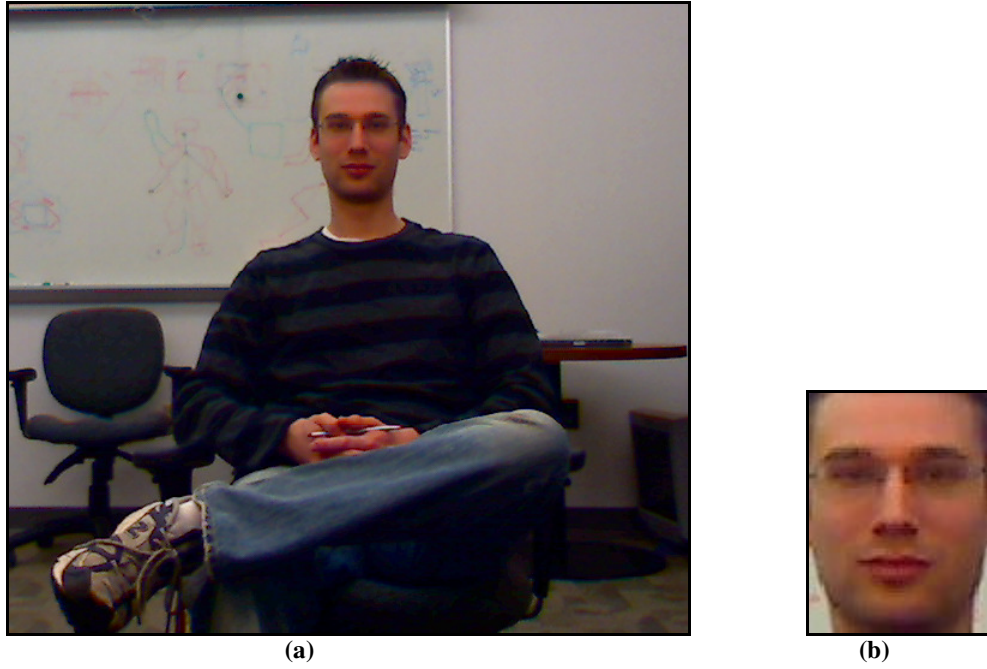
Depending on the number of stages, the creation of the number of positive and negative parameters specified the classification time varies. As a result of the classification process, it generates folders according to the number of stages under “*opencv-1.0.0/apps/haartraining/TrainingFolders/*”. We need to change those folders into xml file format that is needed in face detection program. The utility “*haarconv*” was used to change the folders to the xml file as following,

```
./cascade2xml TrainingFolders/ XMLFileForDetection.xml 25 25
```

where `TrainingFolders` is the collection of the number of folders obtained from training, `XMLFileForDetection.xml` is the output file name of the generation of cascaded folders. This cascade XML file was used in the face detection program written in C.

In order to perform super resolution and recognize faces in video sequences a set of low resolution frames are needed. The face detection system was developed to save the detected face ROI as a JPEG image file. As shown in Figure 15, height and width of the ROI are predetermined depending on the lowest resolution image which could be enhanced by

using super resolution to achieve successful face recognition rates. In our case, 23x30 pixels is the lowest ROI size which could be enhanced to 90x120 to obtain the best face recognition rates. This was determined by evaluation of the super resolution techniques and performance of face recognition rates. In order to obtain a specific size of a frame image, once the face is detected it is cropped to 23x30 or 45x60 to preserve the sizes consistently across the subjects. This is discussed further in Section 4.4.



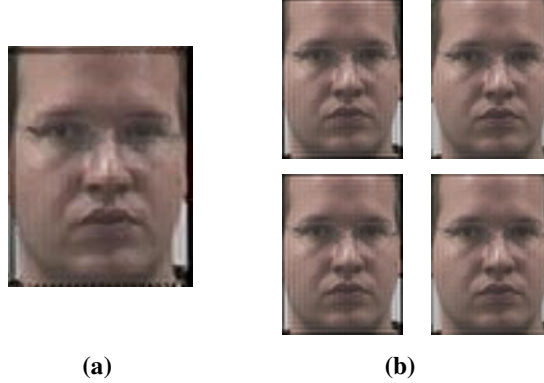
**Fig 15:** (a) Camera captured high resolution frame (b) 90x120 autos cropped interpolated image

## 4.2 Super Resolution

In order to generate a super resolution face image, the program input requires a list of low resolution images to be used. The number of LR images needed varies depending on the resolution to be achieved (which is the size of the high resolution image used for face recognition) and the resolution of the low resolution image. As an example, to obtain 90x120



resolution using 45x60 low resolution frames, which is twice the resolution, four, 45x60 low resolution frames are required. If four times the resolution needs to be achieved, 16, 23x30 low resolution frames/images are needed for the enhancement. Capturing and utilizing more low resolution images does not improve the output of the super resolution [8].



**Fig 16:** (a) 90x120 generated super resolution image (b) four 45x60 low resolution images used to generate 90x120 super resolution image

The next step is to select the SR algorithm, which is done by user interaction with the GUI. As mentioned in Chapter three, two spatial and frequency domain algorithms were implemented and utilized. Following are the steps for the frequency domain super resolution algorithm,

**(1) Compute the Fourier transforms  $f_{LR,m}$  of all low-resolution images.**

**(2) *Rotation estimation:***

Estimate the rotation angle between every image  $f_{LR,m}$  ( $m = 2, \dots, M$ ) and the reference image  $f_{LR,w,1}$ .

**(a)** Compute the polar coordinates  $(r, \theta)$  of the image samples.

**(b)** For every angle  $\alpha$ , compute the average value  $h_m(\alpha)$  of the Fourier coefficients for which  $\alpha-1 < \theta < \alpha+1$  and  $0.1_{\rho} < r < \rho_{\max}$ . The angles are expressed in degrees and  $h_m(\alpha)$  is evaluated every 0.1 degrees. A typical value used for  $\rho_{\max}$  is 0.6.

**(c)** Find the maximum of the correlation between  $h_1(\alpha)$  and  $h_m(\alpha)$  between  $-30$  and  $30$  degrees. This is the estimated rotation angle  $\phi_m$ .

- (d) Rotate image  $f_{LR,m}$  by  $-\varphi_m$  to cancel the rotation.
- (3) **Shift estimation:** The horizontal and vertical shifts between every image  $f_{LR,m}$  ( $m = 2, \dots, M$ ) and the reference image  $f_{LR,1}$  are estimated.
- (a) Compute the phase difference between image  $m$  and the reference image as  $\angle(F_{LR,m}/F_{LR,w,1})$ .
- (b) For all frequencies  $-\mathbf{u}_s + \mathbf{u}_{\max} < \mathbf{u} < \mathbf{u}_s - \mathbf{u}_{\max}$  write the linear equation describing a plane through the computed phase difference with unknown slopes  $\Delta \mathbf{x}$ .
- (c) Find the shift parameters  $\Delta_{xm}$  as the least squares solution of the equations.
- (4) **Image reconstruction:** a high-resolution image  $f_{HR}$  is reconstructed from the registered images  $f_{LR,m}$  ( $m = 1, \dots, M$ ).
- (a) For every image  $f_{LR,m}$ , compute the coordinates of its pixels in the coordinate frame of  $f_{LR,1}$  using the estimated registration parameters.
- (b) From these known samples, interpolate the values on a regular high-resolution grid using bicubic interpolation.

As shown in the above steps, first the Fourier transform of a set of LR images (4 or 16 depending on the final resolution to be achieved) is calculated and a reference LR image is selected randomly. Once the reference image is selected, the rotation angle between every image and the reference image is calculated. Calculation of the rotation angle is done by first computing the polar coordinates  $(r, \alpha)$  of all the Fourier transform of image samples. The average value of Fourier coefficients (AVFC) of every angle  $\alpha$  is computed and the maximum correlation between (-30 and 30 degrees) AVFC of the reference and AVFCs of the other angles were found. This angle is the rotation angle used to cancel the rotation of all the LR images with reference to originally randomly selected LR image. Vertical and horizontal shift estimation was done by computing the phase difference reference LR image and other LR images Fourier transforms. Once shift and rotation estimation is computed, for

every image in the LR image sets new coordinates were calculated and placed in a high resolution grid and bi-cubic interpolation was performed.

In the spatial domain method, assume that the SR image is to be generated from four LR images. First, all the LR images from the same set are assumed, as there is no motion between each LR image. Then motion between the selected LR image 1 and LR image 2 is calculated using the equation,

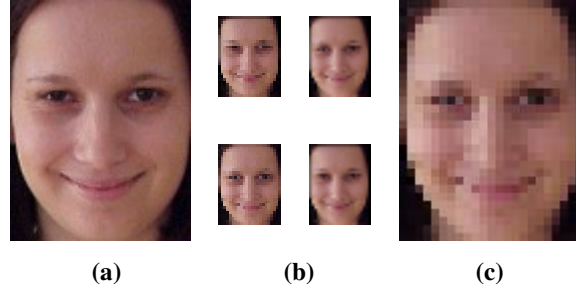
$$\epsilon(\Delta_{x_1}, \Delta_{x_2}, \theta) = \sum \left[ f(x_1, x_2) + \left( \Delta_{x_1} - x_2\theta - x_1 \frac{\theta^2}{2} \right) \frac{\partial f}{\partial x_1} + \left( \Delta_{x_2} + x_1\theta - x_2 \frac{\theta^2}{2} \right) \frac{\partial f}{\partial x_2} \right]^2$$

Once the vertical, horizontal and rotation parameters are calculated, LR image 2 is adjusted and used as the reference image for the LR image 3. This process is followed until all the LR images have zero motion.

### 4.3 Low Resolution Image Generator

A low resolution image generator was implemented as a simulation tool which enabled generating low resolution images from a given high resolution image. The goal of the simulation was twofold. First is to evaluate the performance in a controlled environment with exact knowledge about the original high resolution image and about the shift and rotation values. Second was the sheer difficulty in finding video databases with low and high resolution video sequences. This enabled evaluation of the registration and reconstruction errors. For the simulation, we started from an original high resolution (90x120 pixels) image, these HR images then filtered using an ideal low-pass filter with cutoff frequency  $f_c = 0.125f_{s, \text{original}} - \epsilon$ , with  $f_{s, \text{original}}$  the original sampling frequency of the high resolution image, and  $\epsilon = 0.005f_{s, \text{original}}$ . Next, the filtered image is down sampled by a factor of eight, keeping only every eighth sample. Note that this is different from a normal lower-resolution image, where a set of pixel values from the high resolution image would normally be averaged form a pixel

of the low resolution image. Down sampled images are derived by shifting the set of selected pixels each time by a random number of pixels (0-7) in both horizontal and vertical directions. For the shifts, a standard deviation of 5 is used, while the rotation angles have a standard deviation of 0.5. Figure 17 shows result of low resolution image generator,



**Fig 17:** (a) HR 90x120 image (b) four 45x60 generated low resolution images (c) zoomed image of 23x30 low resolution images to 90x120.

#### 4.4 Face Recognition using Eigenfaces

The algorithm for facial recognition using Eigenfaces is described in Figure 11. First, the original images of the training set are transformed into a set of Eigenfaces  $E$ . Afterwards; the weights are calculated for each image of the training set and stored in the set  $W$ . Upon observing an unknown image  $X$ , the weights are calculated for that particular image and stored in the vector  $W_X$ . Finally,  $W_X$  is compared with the weights of known facial images (the weights of the training set  $W$ ). One way to do the comparison would be to regard each weight vector as a point in space and calculate an average distance  $D$  between the weight vectors from  $W_X$  and the weight vector of the unknown image  $W_X$ . If the average distance exceeds a threshold value, then the weight vector of the unknown image  $W_X$  lies too "far apart" from the weights of the faces. In this case, the unknown  $X$  is considered a non-face. Otherwise (if  $X$  is actually a face), its weight vector  $W_X$  is stored for later classification. The

optimal threshold value was determined empirically. Step by step process for determination of Eigenfaces using PCA is presented below,

**1. Prepare the data:** The faces constituting the training set ( $\Gamma_i$ ), which is set of randomly selected and auto cropped face images.

**2. Subtract the mean:** The average matrix  $\Psi$  has to be calculated, and then subtracted from the original faces ( $\Gamma_i$ ) and the result stored in the variable  $\Gamma_i$ .

$$\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n$$

$$\Phi_i = \Gamma_i - \Psi$$

**3. Calculate the eigenvectors and Eigenvalues of the covariance matrix:** The average  $\Psi$  matrix has to be calculated, then subtracted from the original faces ( $\Gamma_i$ ) and the result stored in the variable  $\Phi_i$

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T$$

The covariance matrix has a dimensionality of  $N^2 \times N^2$ , Therefore  $N^2$  Eigenfaces and Eigenvalues. As an example for a  $120 \times 120$  image calculation would increase computation to a  $14400 \times 14400$  matrix and calculate 14400 Eigenfaces. Computationally, this is not very efficient as most of those Eigenfaces are not useful for our task. Turk and Pentland [48] proposed the scheme below to address the problem. Let

$$L = A^T A; L_{n,m} = \Phi_m^T \Phi_n$$

$$u_l = \sum_{k=1}^M v_{lk} \Phi_k; l = 1, \dots, M$$

where  $L$  is a  $M \times M$  matrix,  $v$  are  $M$  eigenvectors of  $L$  and  $u$  are Eigenfaces. Note that the covariance matrix  $C$  is calculated using the formula  $C = AA^T$ . The advantage of this method

is that one has to evaluate only  $M$  numbers and not  $N^2$ . Usually,  $M \ll N^2$  as only a few principal components (Eigenfaces) will be relevant. The amount of calculations to be performed is reduced from the number of pixels ( $N^2 \times N^2$ ) to the number of images in the training set ( $M$ ). In step 4, the associated Eigenvalues allow one to rank the Eigenfaces according to their usefulness. Typically, only a subset of  $M$  Eigenfaces is used, the  $M'$  Eigenfaces with the largest Eigenvalues.

**4. Select the principal components:** From  $M$  eigenvectors (Eigenfaces)  $u_i$ , only  $M'$  should be chosen, which have the highest Eigenvalues. The higher the Eigenvalue, the more characteristic features of a face does the particular eigenvector describe. Eigenfaces with low Eigenvalues can be omitted, as they explain only a small part of characteristic features of the faces. After  $M'$  Eigenfaces  $u_i$  are determined, the “training” phase of the algorithm is finished.

#### ***4.4.1 Utilizing Eigenfaces***

As shown in Figure 18 first few Eigenfaces are relatively clear as expected. The features seem to match up quite well and a distinct set of eyes, a nose, and a mouth, are clearly visible in each of them. Lack of strange artifacts or noise in the corners of the images is an important feature to note in these Eigenface images.

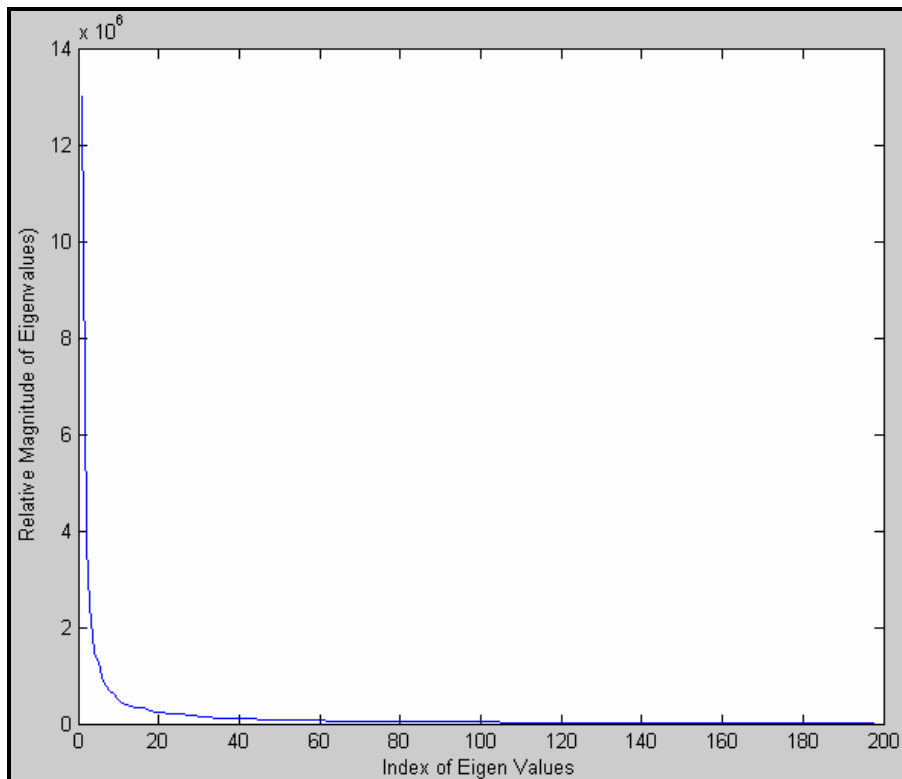


**Fig 18:** First five Eigenfaces

The corners of the image contain background image and are not important facial feature data. As the order of the Eigenfaces increases, and their respective eigenvalues decrease, the images become noisier. This progression is expected because the size of the eigenvalue determines the relevance of the corresponding Eigenface. Therefore as the eigenvalue decreases, the corresponding Eigenface should be capturing less significant components of the face image. This progression is shown below in



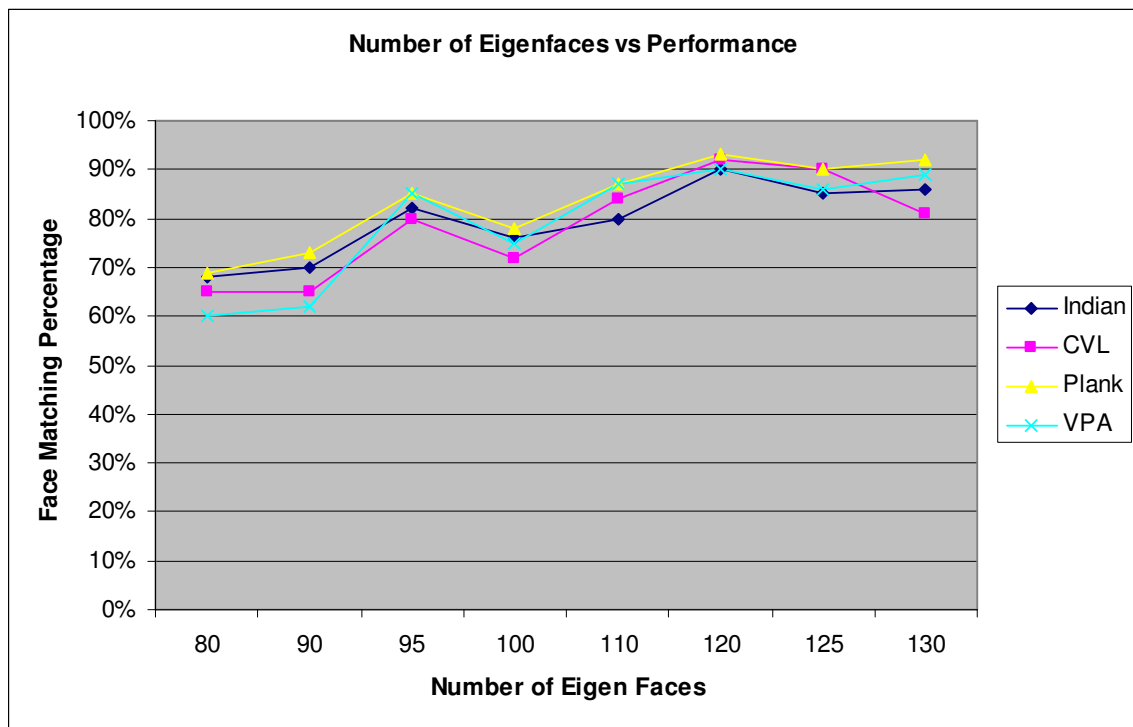
**Fig 19:** Eigenfaces 60 & 80, 100 & 101, and 196



**Fig 20:** Relative magnitude of Eigenvalues

The 60<sup>th</sup> and 80<sup>th</sup> Eigenfaces are still relatively clear face images, but the clarity rapidly continues to degrade until the 196<sup>th</sup> Eigenface only hints at facial features. It is clear from Figure 19 where images contain mostly noise.

One of the important parameters is to determine the number of Eigenfaces to be utilized for face identification. When observing the Figure 20 graph, it is clear that the number of Eigenfaces to be used is in the range of 80-140. Therefore analysis on all four databases was done to find the number of Eigenvalues to be used for optimal results by varying the number of Eigenfaces.



**Fig 21:** Number of Eigenfaces vs. face matching performance

Figure 21 shows that increasing the number of Eigenfaces results in increased performance until it reaches approximately 115 to 120. At this point there is little benefit from the increased number of Eigenfaces. After 120, performance actually begins to decline. These



results were expected, as Eigenfaces themselves degrade into what appears to be nearly useless noise. Thus, Eigenfaces associated with small Eigenvalues are actually capturing noise in the image array and have negative effect on both face recognition performance and speed performance. Another observation is that when the number of Eigenfaces increases, the speed of recognition declines. Therefore 120 Eigenfaces were utilized in the face recognition process.

## **4.5 Limitations and Difficulties**

One of the major difficulties was to get proper sub-pixel level movement in face frames when capturing videos through the camera. Sub-pixel level movement in frames is highly important to have an accurate registration in the super resolution algorithm. Failure in proper registration will have ripple effects in the performance of both generating super resolution images and Eigenface recognition. In order to address this issue, after the face was detected in the first frame, cropping of the ROI was done at the same pixel locations in sequential set of frames assuming there is subpixel level movement.

There are many difficulties inherent in developing a successful set of Eigenfaces. The first and foremost among them is the need for clear, pronounced, relevant Eigenfaces. The only way to achieve such results is to use a large number of faces which have been cropped and rotated in the same manner so that every principal component of the different face images line up with each other. Unfortunately, it is nearly impossible to crop each image by a face detection system in exactly the same manner. To make matters worse the pitch and yaw of the face in the images can not be controlled. While these slight deviations do not matter to the human eye, they have unknown and possibly adverse effects upon the creation of the Eigenfaces.

Each error in cropping or poor image selection potentially introduces an error into the development of the Eigenfaces. If enough error is introduced, an entire Eigenface could be devoted to it, reducing the effectiveness of the other Eigenfaces and increasing the usage of Eigenfaces. Unfortunately, even a visual examination of the Eigenfaces cannot reveal if one is valid or not. To make matters worse, when the matching procedure fails to match a set of faces, it is often extremely difficult to determine why matching failed.

One of the limitations in the face detection portion of the system is distance of the subject and the video camera to obtain a low resolution. When the distance is larger, face detection fails even though the captured frames resolution is higher than the 23x30 resolution. As mentioned in Chapter 3, face resolution of 23x30 was determined as the lowest resolution that a super resolution algorithm could use to generate a frame which could accurately attain face matching.

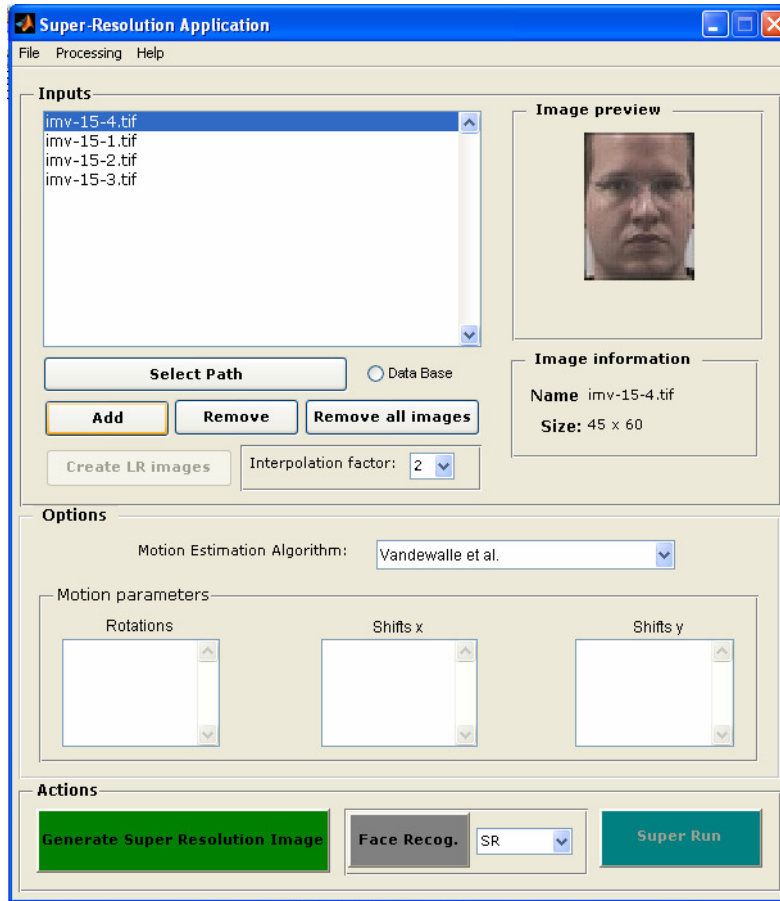
There are many limitations placed upon the effectiveness of the Eigenfaces depending upon their desired use. For the face recognition system, the Eigenfaces are being designed for the identification of the principal components of a subjects face and not simply for identifying faces. This means that the Eigenfaces are always assuming a frontal and non-rotated view of the face and cannot work effectively with a side view of the face. Eigenfaces are generally trained with tilted faces as well as frontal views to allow them to recognize faces that are tilted, as face objects. However, in this work it is assumed that the face will already be identified before it is passed to the Eigenspace for calculations. It is also desired that the pose angle of the face should not play an important factor in identifying whether or not the face matches a particular subject. This line of reasoning clearly leads to the fact that in training the Eigenfaces, it is desired that none of principal components in the Eigenspace correspond to the angle of the face. If too many of the Eigenfaces correspond to the angle of

the face instead of the facial features, the system will match subjects whose heads are posed in the same manner.

Another limitation of the algorithm is that it is significantly more dependant upon lighting variation. While this is not surprising in retrospect, it is still a difficult issue to address. The research showed that an effective Eigenspace was relatively independent of the light source, as long as there was enough light to view the face. Also, the angle of the light will affect the face in varying ways. The amount of highlight on the cheekbones, the amount of shadow on the eyes, and much more depends on the intensity and angle of the light. This means that the environment needs to be more controlled than initially believed if performance is to be maintained. It also may have an effect on how the algorithm generalizes across different environments.

## **4.6 Super Resolution GUI**

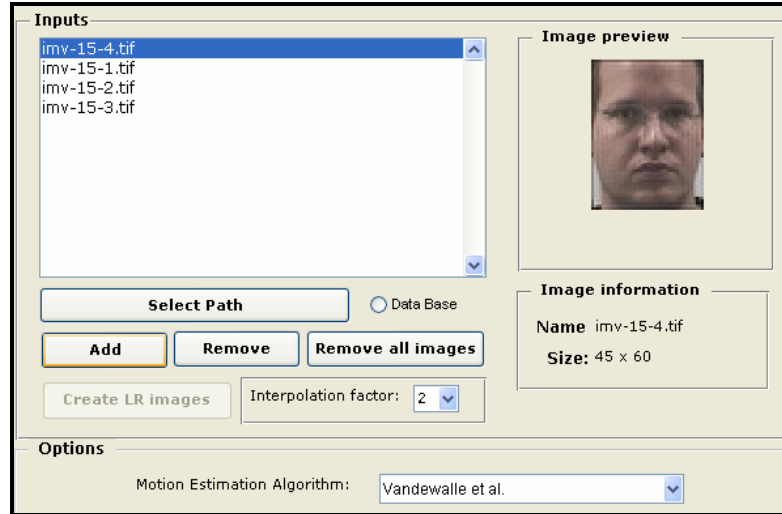
The super resolution graphical user interface (GUI) is an important component in the system even though it performs almost none of the work of the program. The reason for this is that the GUI is the only real method for communicating, controlling, and understanding what the program is doing. Without a user interface, the program is no longer interactive and can no longer inform the user when something unexpected occurs. The layout for the GUI follows a natural progression of data being processed by the system components. A standard left to right, top to bottom succession of components was used in order to make the layout intuitive. A screenshot of the GUI is shown below in Figure 22.



**Fig 22:** Super resolution GUI

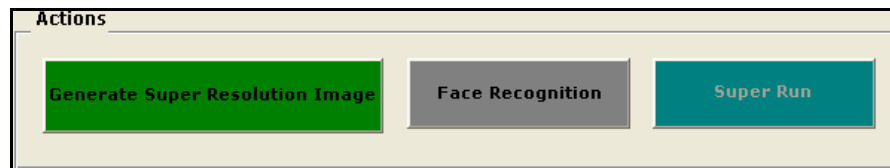
The display and controls for each component are grouped together on the GUI to allow easy understanding and to facilitate a short learning curve. GUI could be separated to two sections. As shown in Figure 22, the input section interacts with the user to the data base path to obtain the input data. The user could use the “*select path*” button to select the location of low resolution images which would be used to generate the super resolution images. The user has the option to process the whole database of low resolution images simply by selecting the “*Data Base*” radio button option. When processing the database, the system processes sets of low resolution images for a subject at a time. The number of low resolution images to be processed depends on the selected interpolation factor (default interpolation factor is set to 2). As an example in Figure 23, since the selected interpolation factor is two, 4 LR images are

selected in the display box. If the interpolation factor is four, 16 low resolution images will be selected. SR algorithm selection menu permits the user to toggle between two different algorithms, which are namely Vandewalle et al. (frequency domain) and Keren et al. (spatial domain) to generate super resolution output.



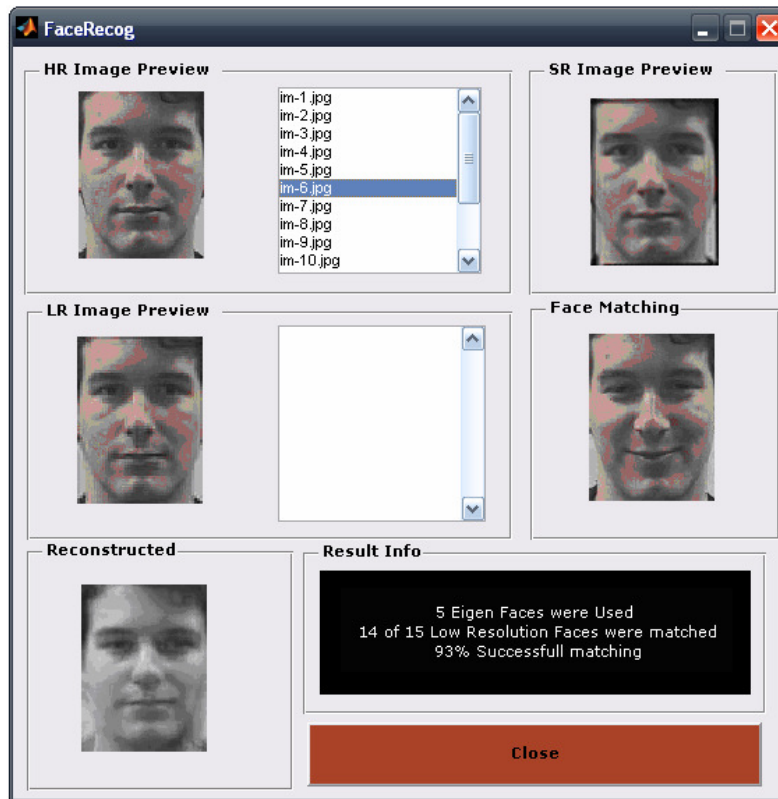
**Fig 23:** Input section of the super resolution GUI

As shown in Figure 24, three actions could be performed using the action section of the super resolution GUI. The buttons named “*Generate Super Resolution Image*” and “*Face Recognition*” are self explanatory.



**Fig 24:** Actions section of the super resolution GUI

Once the face recognition is performed, results are presented using a results GUI. As seen in Figure 25, the user could select the high resolution image from the “*HR Image Preview*” section, which enables display of the respective LR, SR, Face matching and reconstructed Eigenface images.



**Fig 25:** Results GUI

The results info box displays percentage of successful matching and which resolution images were used for the face matching.

## **Chapter 5    Evaluation**

The system is configured to work in two phases. First, the face detection system will be activated and after face detection, an automated cropping utility will generate low or high resolution images from the captured frames depending on the distance of the camera and subject or using the high resolution image/video database. Once that is complete, the results will be used in the super resolution and face matching system which uses the PCA algorithm to determine if the face in the frame is a match for the desired match target. After this match attempt is complete the system will display the results via the results GUI.

### **5.1    Video Database Test**

First tests were done on two video databases mentioned in Chapter 3. The initial test was done using a video database of fifteen subjects. Each high resolution video sequence contains a subject with sub pixel level movements in consecutive frames, directly looking at the camera and was used as an input for the face detection system. The system uses this input to create three sets of data. First, it outputs a cropped 90x120 pixel HR frame. Next, one HR frame is randomly selected by the system and cropped to 90x120 then equalized and saved in the training set as a training image for the face matching. Finally, series of 16 consecutive frames will be cropped to 23x30. Another consecutive four frames will be cropped to 45x60. All frames which are in two sizes 23x30 and 45x60 were then saved as LR images. This process followed for each of the subject and then a database of low resolution faces was created. The identity of the subject was recorded as numbers in the filename so that the system could keep

track of the subject's name during execution. During the first run of the test shown in Table 1a and 1b, face matching results were inadequate for the HR frames.

<b>HR (90x120)</b>	<b>LR (23x30) (Zoomed to 90x120)</b>	<b>SR (90x120)</b>
85%	70%	73%

**Table 2a** - Plank video database first test results using LR 23x30 frames

<b>HR (90x120)</b>	<b>LR (45x60) (Zoomed to 90x120)</b>	<b>SR (90x120)</b>
85%	72%	77%

**Table 1b** - Plank video database first test results using LR 45x60 frames

Low performance results for the HR frames were due to inaccuracies in the registration and motion estimation of the super resolution step. Erroneous registration was due to face detection and automated cropping. Face cropping was executed on an individual frame basis and, if the face was detected, cropping of the face area was done as part of the face detection. Face cropping on each frame introduces more than sub pixel level movement on the cropped frames. This issue was addressed by using a simple technique which is to detect the face on the first frame and use the same location as crop limits on all following frames. This approach yielded better results only if the movement of the subject in consecutive frames is at the sub pixel level, which is a limitation. This modification was followed for the rest of the tests.

The second run of the Plank video database test results were promising. As shown in Figure 26 and in Table 2a and 2b, results show that the system was able to match the face from the HR video sequences to the training set at a rate of 93%. The system matched faces in LR frames, at 80% percentage accuracy for both 23x30 LR dataset and 45x60 LR dataset for frequency domain. Spatial domain method generated 73% and 80% respectively for



23x30 LR dataset and 45x60 LR. Similar but little improvement was shown in interpolated images.

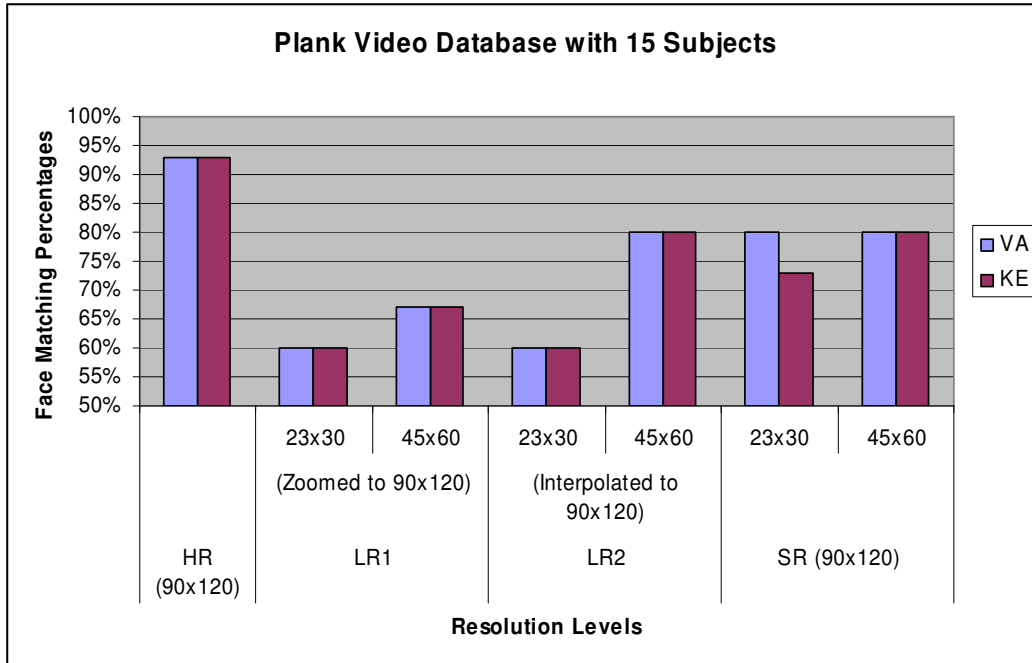
	<b>HR (90x120)</b>	<b>LR1 (23x30) (Zoomed to 90x120)</b>	<b>LR2 (23x30) (Interpolated to 90x120)</b>	<b>SR (90x120)</b>
<b>VA</b>	93%	60%	60%	80%
<b>KE</b>				73%

**Table 2a** –Plank video database results using LR 23x30 frames

	<b>HR (90x120)</b>	<b>LR1 (45x60) (Zoomed to 90x120)</b>	<b>LR2 (45x60) (Interpolated to 90x120)</b>	<b>SR (90x120)</b>
<b>VA</b>	93%	67%	80%	80%
<b>KE</b>				80%

**Table 2b** – Plank video database results using LR 45x60 frames

The performance for both sets of SR images was superior when comparing to that of zoomed and interpolated LR images. Note that the interpolated version is lower in resolution with reference to the HR image, and was referred as a second set of LR images which was compared against the SR results. The results clearly show that SR reconstruction of LR images improves the face recognition rates. Even though this video database contained fewer subjects, this was a good initial test to verify expected system behavior and results. One of the observations from the results would be face recognition rates on SR frames in the case of 23x30 LR frames showed better improvement from both zoomed and interpolated LR frames. On the other hand interpolated version of the LR frames against zoomed version showed no improvement. Therefore, it is evident that interpolated version holds less details and information when compared to SR version, which results in lower performance.



**Fig 26:** Face matching results of video database with 15 subjects (VA [8], KE [9]).

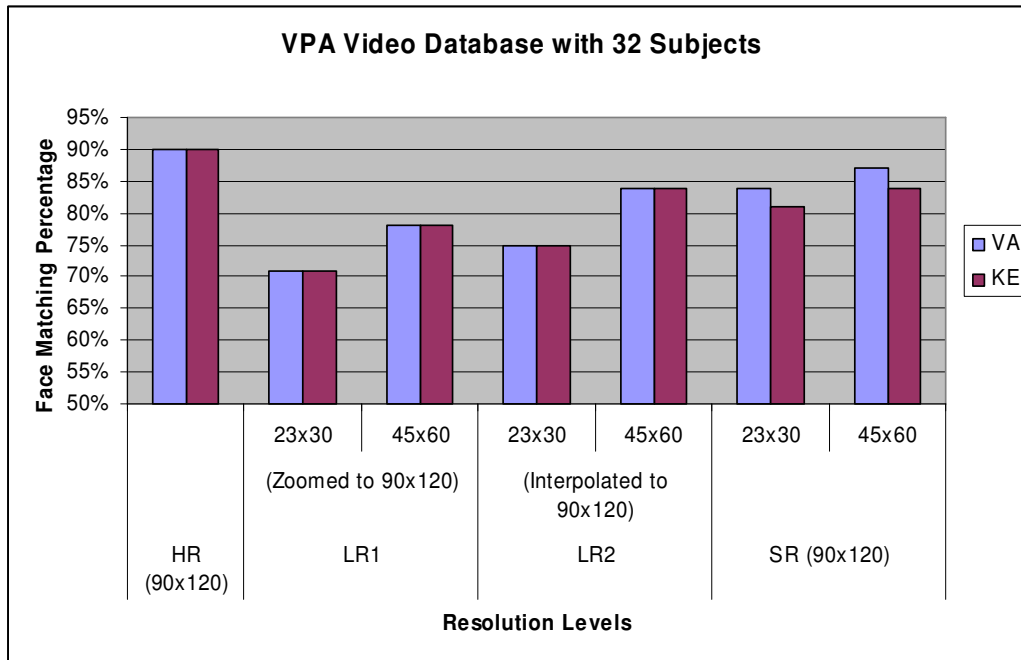
The second test was run using VPA video database. As mentioned in Plank video database testing, first the videos were used as an input for the face detection and cropping system which generated HR, LR(zoomed), LR(interpolated) and SR datasets. These datasets were then used to test face recognition rates against the trained images.

	HR (90x120)	LR1 (23x30) (Zoomed to 90x120)	LR2 (23x30) (Interpolated to 90x120)	SR (90x120)
VA	90%	71%	75%	84%
KE				81%

**Table 3a** – VPA video database results using LR 23x30 frames

	HR (90x120)	LR1 (45x60) (Zoomed to 90x120)	LR2 (45x60) (Interpolated to 90x120)	SR (90x120)
VA	90%	78%	84%	87%
KE				84%

**Table 3b** – VPA video database results using LR 45x60 frames



**Fig 27:** Face matching results of VPA video database with 32 subjects (VA [8], KE [9]).

The data in Table 3a and 3b can be read as follows. Face recognition percentage accuracy rates of 90% were attained for 90x120 HR video frame set. LR image sets 23x30 and 45x60 zoomed to 90x120 respectively recorded 71% and 78% for the recognition rates. The results from Table 3a and 3b show that the system performed just as accurately with a testing database of thirty two subjects as it did in the initial test of fifteen subjects. Higher percentage of face matching accuracy in SR frames comparing to LR frames clearly shows in both frequency and spatial approaches to super resolution could use as a performance booster for face recognition in LR frames/images.

## 5.2 Image Database Test

It should be noted that the number of subjects in the video database was small and it is a real challenge to find databases with color video that are suitable for the needs of this research. On the other hand, it is trouble-free to find color image databases, where lot of research has been done. Hence a simulated approach was taken by utilizing image databases. Each high resolution image was considered as a video frame. Images were obtained from two well known image databases, namely The Indian face database [44] and CVL face database [45]

First the Indian face database [44], which included 60 distinct subjects with eleven different poses for each individual, was tested. A total of 55 subjects were hand selected from the database to have minimum head yaw. Out of the eleven different poses, two were selected where one was selected as the HR image and one was selected for training. As shown in Figure 28, this database had distinct numbering for different poses. First, by observing the database two poses were selected where the subjects directly looking at the camera with a minimum yaw. Once the image selection was done, the face detection and cropping system was modified to accept a file list with selected image names which uses for image reading purpose. This creates a dataset of cropped 90x120 HR images.



**Fig 28:** One subject from the Indian face database with two different poses. (a) selected as HR image and (b) selected as training image for the face recognition system.

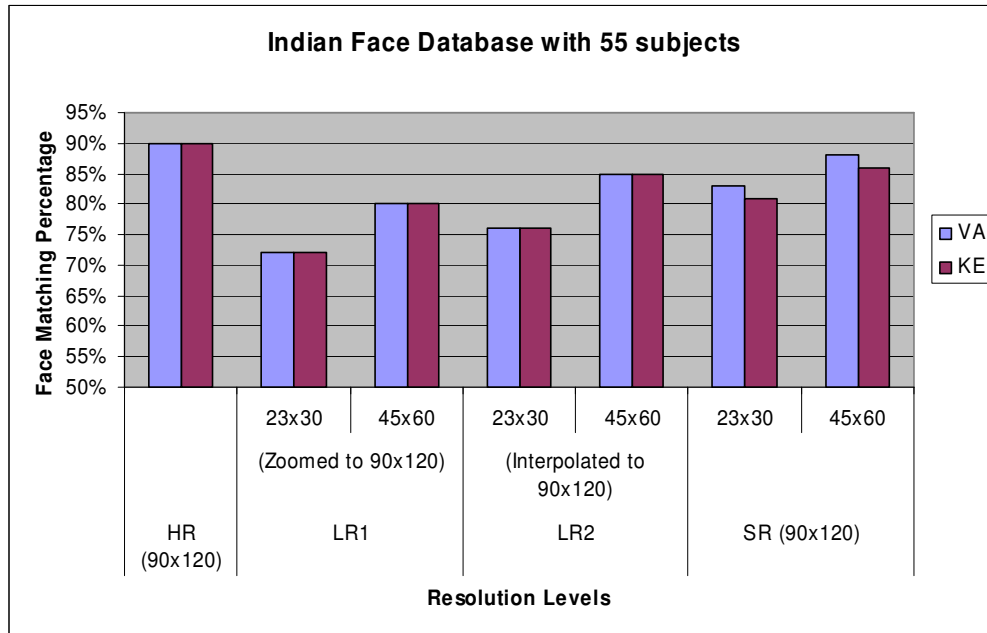
In the next step, 23x30 and 45x60 LR images were generated using the low resolution image generator. The interpolated version of one LR image with size 90x120 per subject was created to compare with SR results. Results shown in Table 4a and 4b were observed for face recognition using two super resolution algorithms.

	HR (90x120)	LR1 (23x30) (Zoomed to 90x120)	LR2 (23x30) (Interpolated to 90x120)	SR (90x120)
VA	90%	72%	76%	83%
KE				81%

**Table 4a** – The Indian face database results using LR 23x30 images

	HR (90x120)	LR1 (45x60) (Zoomed to 90x120)	LR2 (45x60) (Interpolated to 90x120)	SR (90x120)
VA	90%	80%	85%	88%
KE				86%

**Table 4b** – The Indian face database results using LR 45x60 images



**Fig 29:** Face matching results of Indian face database with 55 subjects (VA [8], KE [9]).

As expected, an increase in the number of subjects was accompanied by an increase in number of false positive matches. However, the increase in correct matches outweighed the increase in false positives. This data corresponds closely with what was anticipated from the early PCA testing. A face should fall into the category of unmatched if the image is cropped poorly or training subjects' faces are not lining up in Eigenfaces.

This theory was proven true through examination of the captured images. The images which remained unmatched were mostly faces which had slight rotation, as in Figure 25. The most important and objective of the test is to verify that SR techniques enhance the face matching results. This was easily proven from the results shows in Figure 24 where,

$$\mathbf{HR}_{percentage\ accuracy} > \mathbf{SR}_{percentage\ accuracy} > \mathbf{LR}_{percentage\ accuracy}$$



**Fig 30:** Examples of unmatched images

A second image database tested was CVL face database which featured 107 persons with seven poses of the each person. A total of 97 subjects were hand selected to reduce the number of images with rotated faces. As in the first test, similar steps were taken to generate the HR, LR(zoomed), LR(interpolated) and SR data sets. The CVL database test results verified that with a large number of subjects, the expected results were achieved and

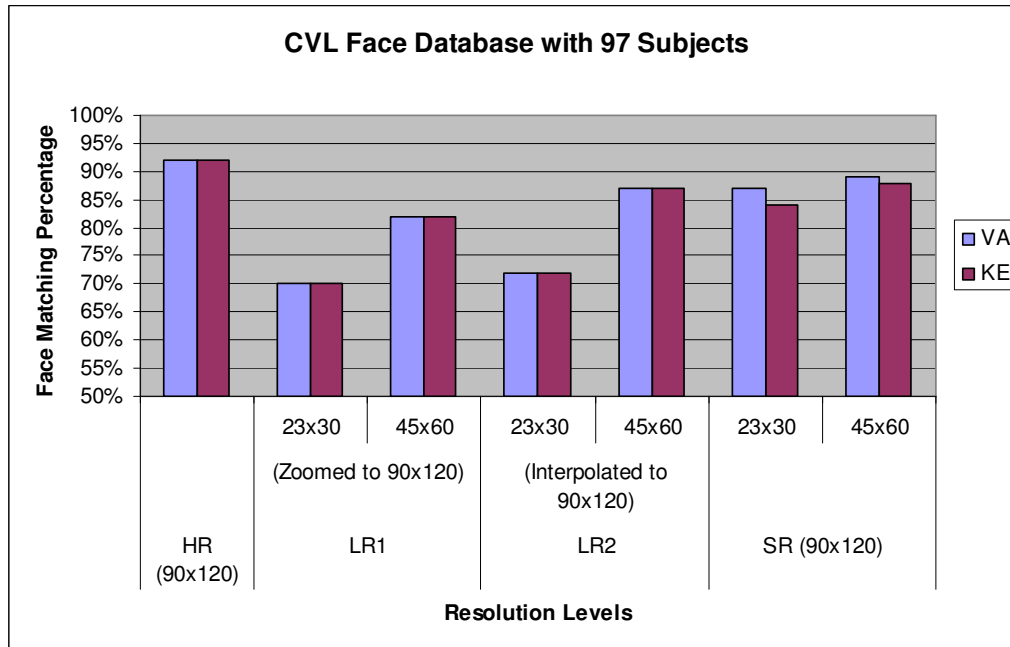
$$\mathbf{HR}_{percentage\ accuracy} > \mathbf{SR}_{percentage\ accuracy} > \mathbf{LR}_{percentage\ accuracy} \text{ is achievable.}$$

	HR (90x120)	LR1 (23x30) (Zoomed to 90x120)	LR2 (23x30) (Interpolated to 90x120)	SR (90x120)
VA	92%	70%	72%	87%
KE				84%

**Table 5a** - CVL image database results using LR 23x30 images

	HR (90x120)	LR1 (45x60) (Zoomed to 90x120)		LR2 (45x60) (Interpolated to 90x120)		SR (90x120)
VA	92%	-	82%	-	87%	89%
KE						88%

**Table 5b** - CVL image database results using LR 45x60 images



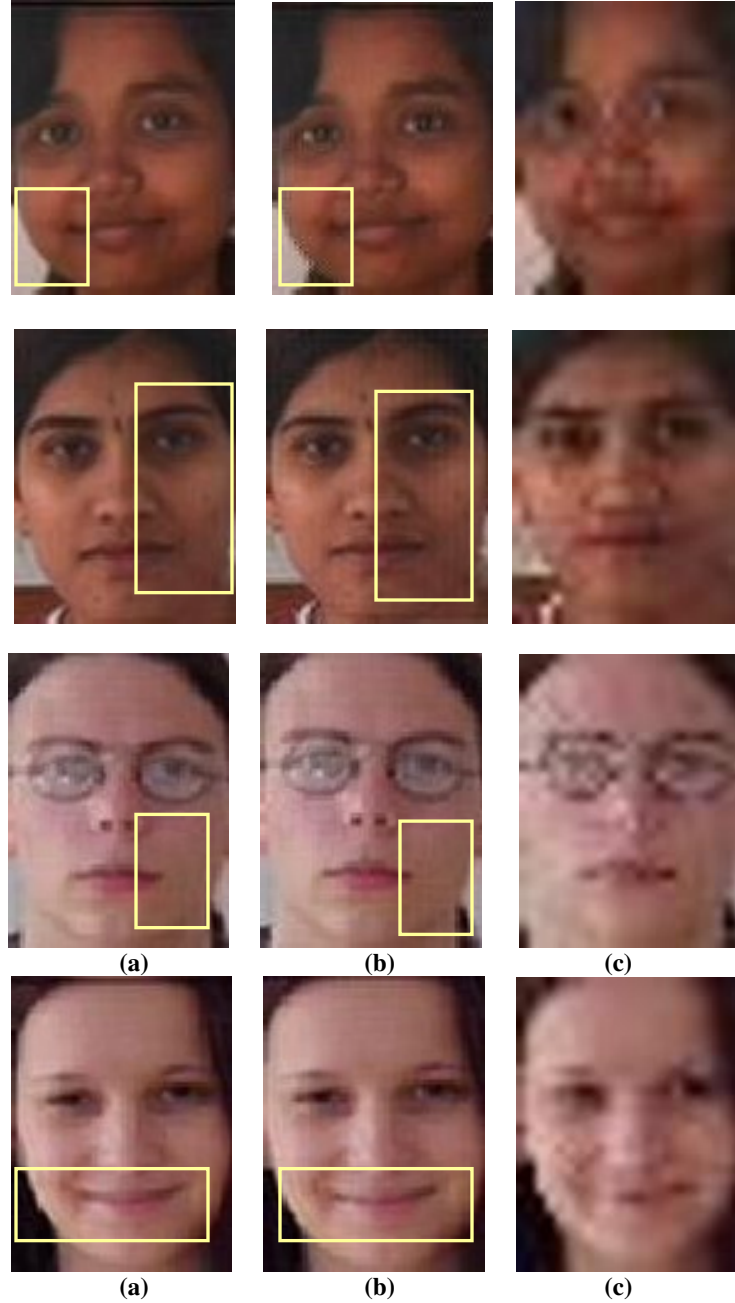
**Fig 31:** Face matching results of CVL face database with 97 subjects (VA [8], KE [9]).

Significant observation in results for all image & video databases was that, for facial recognition accuracy, even though spatial domain and frequency domain performances were close, the frequency domain algorithm outperformed the spatial domain method. The frequency domain method performs well over spatial domain method by Karen *et al.* when

the edge content of an image is high. The accuracy of rotation estimation and shift estimation depends on the presence of some edges in the images [8]. The dependency of the edges is due to the use of frequency domain for motion estimation. Face images contain smooth edges which results in better rotation and shift estimations consequently result in higher super resolution face recognition rates in frequency domain.

Even when analyzing the image results from both image databases it is clear that the frequency domain algorithm performed better compared to spatial domain algorithm and bicubic interpolation. As shown in Figure 27, the frequency domain algorithm results are much smoother with fewer artifacts. Highlighted areas in Figure 27 clearly show the discrepancy where more in spatial domain results. Figure 27 also clearly shows that bicubic interpolation result is degraded compared to outputs from SR methods. This clearly shows that merely up sampling and interpolating does not add the missing frequency content needed for recognition.





**Fig 32:** SR results from two algorithms with areas selected (a) results from VA algorithm (b) results from KE algorithm (c) results from bi-cubic interpolation

While comparing the results to different methods available in the literature, Sezer, Altunbasak and Ercil introduced a face subspace approach in [33]. Osman has used a cosine similarity metric shown below to capture the face recognition.

$$d = \frac{a_{training} \bullet a_{test}}{\|a_{training}\| * \|a_{testing}\|}$$

Table 6 shows summary of results Osman gathered using cosine similarity metric.

	1. Bilinear interpolation	2. SR space domain	3. MAP Estimation	4. POCS w/ outliers	5. POCS w/ variance
<b>PCA</b>	56.25	56.25	75	62.5	62.5
<b>ICA</b>	56.25	56.25	75	62.5	62.5

**Table 6** - Recognition performance (%) using cosine similarity metric for: Column # 1: Bilinear interpolation, Column # 2: Pixel-domain super resolution, Column # 3: Subspace-based Bayesian (MAP) estimation (the result in the ICA row is the result of the proposed method), Column # 4: Subspace-based POCS with outliers of residual constraint, Column # 5: Subspace-based POCS with variance of the residual constraint.

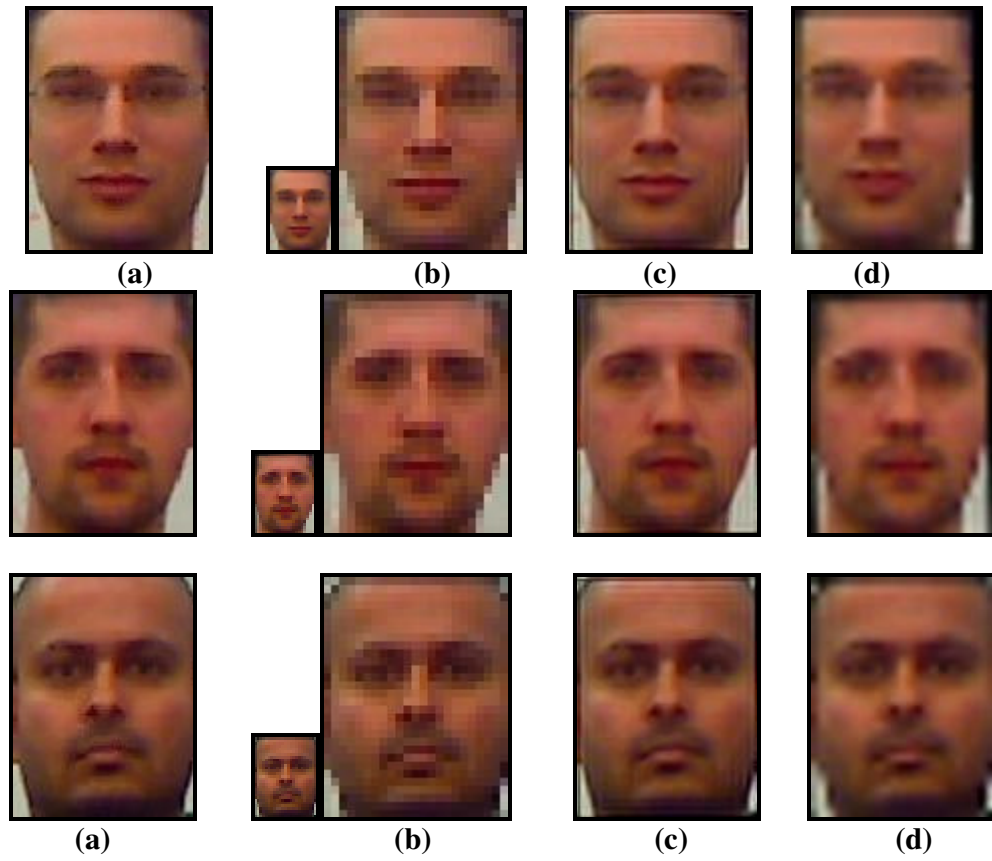
Osman reports that the pixel domain SR algorithm works well when the noise level is low and motion estimation errors are small. It fails to improve the recognition rate compared with the straight forward application of bilinear interpolation shown in the first column. However, different results were seen when comparing frequency domain method and spatial domain methods to bilinear interpolation results. All the results indicated that bilinear interpolation results were lower comparing to the super resolution techniques exploit in this thesis.

### 5.3 Web Camera Captured Video Testing

Web camera performance testing was done using a database of six subjects. The six subjects were gathered and asked to stand in front of the camera system, one person at a time. Multiple images of each subject's face were captured with the Logitech web camera [49]. These images were cropped using the automatic image cropping tool and examined to determine which image was the best face image. The best face image would be one in which the subject was directly facing the camera, the head was not rotated along any axis, hair was

not obscuring the face, and the eyes were open. This process was repeated for each of the six subjects and then a small database of HR faces was created and one more HR image per subject with a different pose was added to the training set increasing the training set to 204 images. The identity of the subject was recorded in the filename so that the system could keep track of the subjects name during execution.

In order to accurately test each individual, only one person was allowed to be in the cameras field of view at a time. This rule ensured that every face which was found by the camera system during the test run was from the same subject. This would allow both true positive matches and false positive matches to be determined from the results. Each individual was asked to remain in front of the camera system for a couple of minutes in two different resolution settings of the camera to obtain HR and LR frames. In both times, the system logged approximately 30 frames worth of data. As in Figure 28, first using the HR setting, the system creates two 90x120 HR images and adds one to the training database and one to the HR testing database. Secondly the system saves four image frames in size 23x30 and 16 frames of 45x60 frames when the camera is in LR settings, which were used as LR frames. In order to ensure that the subject was able to change facial pose and location without seriously affecting the matching potential, the subject was asked to sit still with subtle movements. Rapid movement was discouraged to reduce the movements more than sub pixel levels. The results of the analysis are shown below in Figure 29.



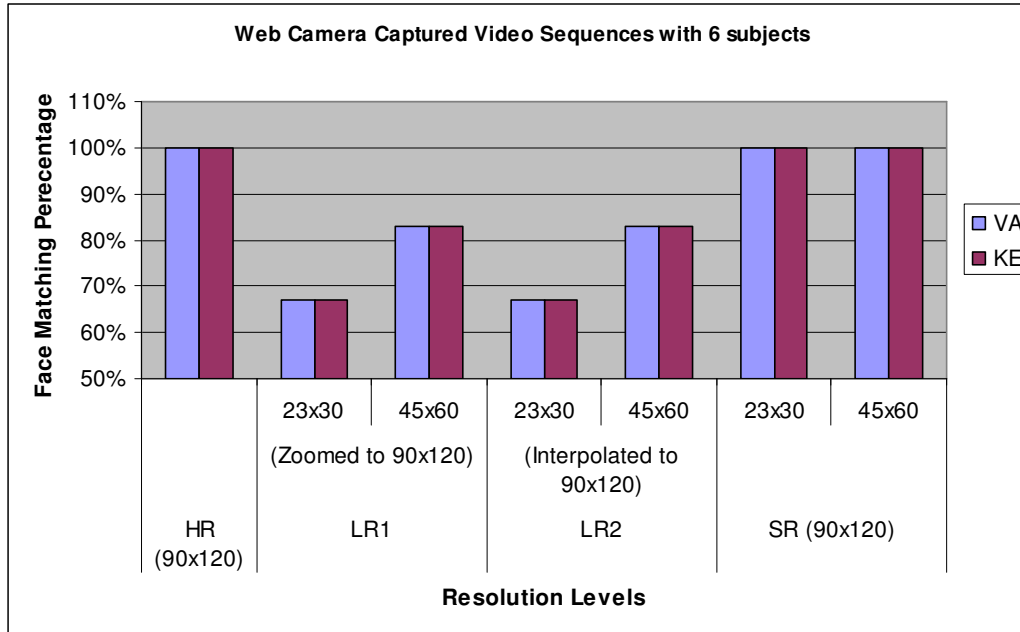
**Fig 33:** Captured & auto cropped frames from Logitech web camera (a) 90x120 HR frame (b) 23x30 LR frame & 90x120 zoomed version (c) 90x120 reconstructed super resolution image using 16, 23x30 frames. (d) bicubic interpolation of a 23x30 frame

	HR (90x120)	LR1 (23x30) (Zoomed to 90x120)	LR2 (23x30) (Interpolated to 90x120)	SR (90x120)
VA	100%	67%	67%	100%
KE				100%

**Table 7a** – Face recognition results of web camera captured video for 23x30 frames

	HR (90x120)	LR1 (45x60) (Zoomed to 90x120)	LR2 (45x60) (Interpolated to 90x120)	SR (90x120)
VA	100%	83%	83%	100%
KE				100%

**Table 7b** – Face recognition results of web camera captured video for 45x60 frames



**Fig 34:** Face matching results of web camera captured video with 6 subjects (VA [8], KE [9]).

The data in Table 7a and 7b can be read as follows. Out of the six subjects, five and four subjects were correctly matched with the trained subjects to the LR frames with respective resolutions 23x30 and 45x60. Both SR and HR frames performed perfectly with 100% face matching accuracy. When analyzing LR results in Table 7b against Table 7a we could see that there is a spike in accuracy rates in comparison to results obtained from image and video databases for LR images/frames. This spike is an artificial one due to the small number of subjects. The LR frames of size 23x60 failed in recognizing only one image more than the LR frames of size 45x60. This is even clearer from the SR results where it is at 100%. Even though a smaller number of subjects were used in this test, still it could be derived that there is a positive effect when using SR enhancement in face matching. Interpolation of the LR frames clearly did not have a profound effect on the recognition level. Interpolation output in Figure 20(d) shows a blur effect on the image indicating the lack of high frequency content.

## 5.4 Timing Performance

Task	Avg. Time
<b>Face Detection and Auto Image cropping</b>	0.85 s
<b>Low Resolution Image Generation</b>	
<i>23x30, 16 Images</i>	7.10 s
<i>45x60, 4 Images</i>	1.28 s
<b>Super Resolution Image Reconstruction</b>	
Frequency Domain	
<i>23x30, 16 Images</i>	10.63 s
<i>45x60, 4 Images</i>	9.03 s
Spatial Domain	
<i>23x30, 16 Images</i>	2.26 s
<i>45x60, 4 Images</i>	1.43 s
<b>Face Matching</b>	0.22 s

**Table 8**– System timing performance

Timing performance results are shown in Table 8. As expected, the most significant timing portion of the face matching process is related to the actual SR image reconstruction. It is important to examine the timing performance differences between spatial and frequency domain approaches for SR reconstruction. The frequency domain method used in this thesis has a higher computational intensity, as all the images need to be registered jointly; they are inherently more complex than the spatial domain method that allow pair wise registration. Therefore, when utilizing the frequency domain approach computational intensity is a trade off which one should be prepared to endure.

Therefore, it is evident that since the spatial domain method has average performance and superior performance in timing it is more suitable for real time video based super resolution applications. On the other hand, frequency domain method could be used in

forensic, scientific, medical and satellite imaging where increased resolving power is important. It is important to explore a method to maintain the same SR improvement levels while not sacrificing speed. In the frequency domain method, the first reference LR image can be used to calculate the motion estimation of the other LR images in the sequence using parallel processing which is not possible in spatial domain. This could be a future improvement for the frequency domain method which could be made to obtain better or matching speed rates compared to the spatial domain method.

## **5.5 System Limitations**

There are limitations on the system that are derived directly from the cameras used. Two of the most important limitations are related to the camera are the level of focus and the possibility of blur in the image. As a subject changes his or her distance from the camera, there is often a chance that the face will be slightly out of focus. Even if it is not substantial, the lack of focus affects the accuracy with which the facial features can be selected. Add this to the possibility that the image will be slightly blurred from subject movement and it is possible that some of the frames are simply not sharp enough to produce an accurate face match. Therefore, a predefined maximum distance from the web camera was found by many test runs and defined when capturing the face frames from camera.

Aside from the hardware limitations, many of the sections of the algorithm are too dependent upon the lighting of the image. A face detection system is generally more prone to shadows. If the lighting is poor, face detection part would fail, causing a failure in the face recognition. However, more robust methods require too much processing and would limit the performance of the system as well.

## Chapter 6 Conclusion

Face recognition has been an active and fascinating field and will be engaging researchers in years to come with diverse challenges and fertile ground for new ideas. It has many diverse uses such as part of a security surveillance system or being in applications that make our life easier or drastically help the disabled individuals lead an independent life. A face recognition system capable of using low resolution images for enhanced performance using super resolution was demonstrated. Two approaches, Vandewalle, et al. [8] and Keren et al. [9], were implemented and compared. Also, as part of the simulation of image database as video database a low resolution image generator with sub-pixel level movement was developed. PCA algorithm with Eigenfaces was implemented for the face matching.

From all the testing and evaluation, the expected performance result was demonstrated:

$$\mathbf{HR}_{percentage\ accuracy} > \mathbf{SR}_{percentage\ accuracy} > \mathbf{LR}_{percentage\ accuracy}$$

which means that improved face recognition rates are obtained using SR to enhance the LR frames/images. Also, the importance of high frequency content in face recognition was evident when SR results were compared against interpolated results. Therefore, simply upsampling does not always enhance the LR images to the desired state where accurate face matching may be obtained. It was clear that the frequency domain approach for face recognition was superior to the spatial domain one whilst the frequency domain approach was slower in timing performance. The PCA algorithm was robust to small variations in head pose and facial expression. Despite the success of the use of PCA, testing indicated that further improving the performance of the matching process may not be possible.



Theoretically, increasing the number of training images would improve the quality of the Eigenfaces, but in practice too much noise is added to the system. In addition, while the early Eigenfaces appear to be the only ones which correspond to significant changes in lighting, other Eigenfaces are affected as well. In order to create a system which can function on images from multiple environments, a more robust face recognition algorithm should be used. If this is not possible, then the system must operate within a constrained environment.

## **6.1 Future Work**

There was degradation in performance for the super resolution reconstruction between images from databases with actual camera captured frames. This is due to automated face cropping which introduced more than sub pixel level movements to the frames. Even though an alternative method was adopted where face detection is performed in the first frame and the same corner location is used for cropping in following frames, a better solution would be to implement eyes and mouth location detectors and use these points to generate crop limits.

Lastly, in order to make the face matching algorithm more robust, it would benefit the system if it were able to handle matching faces with occlusions. While this problem is difficult, many people commonly wear glasses or a baseball cap, and face recognition should be possible under such conditions.

## Bibliography

- [1] V. Bruce, P.J.B. Hancock and A.M. Burton, "Comparisons between human and computer recognition of faces", *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 408-413, Apr 1998.
- [2] T. E. Boult, M.-C. Chiang, and R. J. Micheals, "Super-Resolution via Image Warping," *Super-Resolution Imaging*, Ed. S. Chaudhuri, Boston: Kluwer Academic Publishers, 2002, pp. 131-169.
- [3] S. Baker and T. Kanade, "Hallucinating Faces," *Fourth International Conf. Automatic Face and Gesture Recognition*, March 2000.
- [4] C. Liu, H.-Y. Shum, and C.-S. Zhang. "A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local Nonparametric Model," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, December 2001.
- [5] D. Capel and A. Zisserman, "Super-Resolution from Multiple Views Using Learnt Image Models," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, December 2001.
- [6] B. K. Gunturk, A.U. Batur, Y. Altunbasak, M. H. Hayes III, and R. M. Mersereau, "Eigenface-based Super-resolution for Face Recognition," *Proc. International Conf. on Image Processing*, 2002.
- [7] L. Wiskott, J. M. Fellous, N Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," Tech. Rep. IR-INI 96-08, Institute fur Neuroinformatik, Ruhr-Universitat Bochum, D-44780 Bochum, Germany, 1996.
- [8] P. Vandewalle, S. Ssstrunk and M. Vetterli, "A Frequency Domain Approach to Registration of Aliased Images with Application to Super-Resolution", *EURASIP Journal on Applied Signal Processing (special issue on Super-resolution)*, Vol. 2006, pp. Article ID 71459, 14 pages, 2006.
- [9] D. Keren, S. Peleg and R. Brada, "Image sequence enhancement for super-resolution image sequence enhancement," in *Proceeding IEEE Conference on Computer Vision and Pattern Recognition*, pp. 742-746, June 1988.
- [10] M. Turk and A. Pentland. "Eigenfaces for recognition". *Journal of Cognitive Neuroscience*, vol. 3, pp. 71-86, 1991
- [11] B. K. Gunturk, A.U. Batur, Y. Altunbasak, M. H. Hayes III, and R. M. Mersereau, "Eigenface-based Super-resolution for Face Recognition," *Proc. International Conf. on Image Processing*, 2002.

- [12] T. Kanade. "Picture Processing by Computer Complex and Recognition of Human Faces." PhD thesis, Kyoto University, 1973.
- [13] CV Reference Manual, "*Pattern Recognition: Object Detection*", [http://opencvlibrary.sourceforge.net/CvReference#cv\\_pattern](http://opencvlibrary.sourceforge.net/CvReference#cv_pattern)
- [14] Paul Viola and Michael Jones, "*Rapid Object Detection using a Boosted Cascade of SimpleFeatures*", in Conference on Computer Vision and Pattern Recognition, 2001.
- [15] R. Lienhart and J. Maydt, "An extended Set of Haar-like Features for Rapid Object Detection", *IEEE ICIP 2002*, vol 1, pp. 900-903, Sep.2002.
- [16] V. Bruce, P.J.B. Hancock and A.M. Burton, "Comparisons between human and computer recognition of faces", *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 408-413, Apr 1998.
- [17] G. Yongsheng and M.K.H. Leung, "Face recognition using line edge map", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, Iss.6, pp.764-779, Jun 2002
- [18] J.E. Meng, W. Chen and W. Shiqian, "Highspeed face recognition based on discrete cosine transform and RBF neural networks"; *IEEE Transactions on Neural Networks* vol. 16, Issue 3, pp.679 – 691, May 2005.
- [19] S. Da-Rui and W. Le-Nan, "A local-to-holistic face recognition approach using elastic graph matching" *Proc. International Conference on Machine Learning and Cybernetics*, vol. 1, pp. 240 - 242, 4-5 Nov. 2002.
- [20] W. Haiyuan, Y. Yoshida, T. Shioyama, "Optimal Gabor filters for high speed face identification" *Proc. 16th International Conference on Pattern Recognition*, vol. 1, pp. 107 - 110, Aug 2002.
- [21] R.Y. Tsai and T. S. Huang, "Multiframe image restoration and registration", in *Advance in Computer Vision and Image Processing*. JAI Press, 1984, vol. 1, pp. 317-339.
- [22] S. Park, M. Park, and M. Kang, "Super resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, pp. 21-36, May 2003.
- [23] S. Borman and R. Stevenson, "Spatial resolution enhancement of low-resolution image sequences – a comprehensive review with direction for future research," University of Notre Dame, Tech Rep., 1998. [Online]. Availabe: <http://www.nd.edu/~sborman/publications/SReview.pdf>
- [24] S. Park, M. Park, and M. Kang, "Super resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, pp. 21-36, May 2003.
- [25] S. Chaudhuri, *Super-resolution Imaging*, Kluwer Academic Publishers, 2001.

- [26] S. Kim, N. Bose, and H. Valenzuela, "Recursive reconstruction of high resolution image from noisy under sampled multi-frames", *IEEE Trans. Acous., Speech, Signal Processing*, vol. 38, pp. 1013-1027, June 1990.
- [27] B. Marcel, M. Briot and R. Murrieta, "Calcul de translation et rotation par la transformation de Fourier." *Traitement du Signal*, vol. 14, no. 2, pp 135-149, 1997.
- [28] M. Irani and S. Peleg, "Improve resolution by image registration," *Graph. Models Image Process.*, vol. 53, pp.231-239, March 1991.
- [29] R. Schulz and R. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Procesing.*, vol. 5, pp. 996-1011, June 1996.
- [30] R. Hardie, K. Barnard, and E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of under sampled images," *IEEE Trans. Image Procesing.*, vol. 6, pp. 1621-1633, Dec. 1997.
- [31] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Proceedings of 2nd European Conference on Computer Vision (ECCV '92)*, Lecture Notes in Computer Science, pp. 237-252, Santa Margherita Ligure, Italy, May 1992.
- [32] F. Lin, J.Cook, V. Chandran and S. Sridharan, "Face recognition from super resolved images" *IEEE Trans. Image Procesing*, Volume 12, Issue 5, pp. 597-606, May 2003.
- [33] Osman Gokhan Sezer, Yucel Altunbasak and Aytul Ercil, "Face recognition with independent component based super-resolution," *IEEE Trans. Neural Networks*, Volume 13, Issue 6, pp. 1450-1464, Nov 2002.
- [34] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing*, vol. 5, no. 8, pp. 1266-1271, 1996.
- [35] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison Wesley, 2nd Ed. 2002
- [36] C. Liu, H. Wechsler, "Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition," *Proc. of the Second International Conference on Audio- and Video-based Biometric Person Authentication, AVBPA'99*, pp. 211-216, Washington D.C., USA, March 1999.
- [37] M. Bartlett and T. Sejnowski. "Independent components of face images: A representation for face recognition," *Proc. the 4th Annual Joint Symposium on Neural Computation*, Pasadena, CA, May 17, 1997.

- [38] L. Wiskott, J.-M. Fellous, N. Krueger, C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," Chapter 11 in *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, Eds. L.C. Jain et al., CRC Press, pp. 355-396, 1999.
- [39] B. Moghaddam, T. Jebara, A. Pentland, "Bayesian Face Recognition, Pattern Recognition," Vol. 33, Issue 11, pp. 1771-1782, November 2000.
- [40] C. Liu, H. Wechsler, "A Unified Bayesian Framework for Face Recognition," *Proc. of the 1998 IEEE International Conference on Image Processing*, pp. 151-155, October 1998.
- [41] B. Moghaddam, T. Jebara, A. Pentland, "Bayesian Face Recognition," *Pattern Recognition*, Vol. 33, Issue 11, pp. 1771-1782, November 2000.
- [42] C. Liu, H. Wechsler, "A Unified Bayesian Framework for Face Recognition," *Proc. of the 1998 IEEE International Conference on Image Processing*, pp. 151-155, October 1998.
- [43] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection," *Proc. of the 4th European Conference on Computer Vision*, ECCV'96, Cambridge, UK, pp. 45-58, April 1996.
- [44] Vidit Jain, Amitabha Mukherjee. The Indian Face Database. <http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/> , 2002.
- [45] Peter Peer, CVL Face Database: <http://www.lrv.fri.uni-lj.si/facedb.html>
- [46] VPA Superresolution Face Database: <https://vpa.sabanciuniv.edu/phpBB2/viewtopic.php?t=4284&sid=9a62c4dd1e90800efc344e31f3535faa>
- [47] Open Source Computer Vision Library, <http://www.intel.com/technology/computing/opencv/>
- [48] M. Turk, A. Pentland. "Eigenfaces for Recognition". *Journal of Cognitive Neuroscience*. Vol 3, No. 1, pp. 71-86, 1991
- [49] Logitech Orbit Sphere AF Specs [http://www.logitech.com/index.cfm/webcam\\_communications/webcams/devices/3480&cl=au,en](http://www.logitech.com/index.cfm/webcam_communications/webcams/devices/3480&cl=au,en)